

高等学校电子信息类教材

# 无线数据通信技术

**Radio Technology of Data Traffic**

杜思深 主编

甘忠辉 副主编

郭建新 薛凤凤 林宝勤 蒋磊 等参编

葛建华 主审

電子工業出版社

**Publishing House of Electronics Industry**

北京 • BEIJING



## 内 容 简 介

本书以无线数据通信系统为主线,以军事战术通信为焦点,介绍正在应用和发展的信源与信源压缩编码技术、抗干扰扩频调制通信技术、无线通信多址技术、无线通信组网技术以及无线数据通信协议、通信加密与解密、差错控制与信道编码等新技术,以使读者对无线数据通信领域的新技术、发展趋势有一个基本的了解。在编写过程中力求简明扼要、深入浅出,注重对内容的提炼,避免了抽象的理论表述和复杂的公式推导。

本书可作为现代无线数据通信技术参考书,也可作为一本现代军事战术通信技术的参考书,适合于通信及相关专业的科研人员、教学人员、研究生、高年级本科生以及科研管理人员、装备检验人员、通信技术干部和指挥人员使用。

未经许可,不得以任何方式复制或抄袭本书之部分或全部内容。

版权所有,侵权必究。

## 图书在版编目(CIP)数据

无线数据通信技术/杜思深主编. —北京:电子工业出版社,2011.1

高等学校电子信息类教材

ISBN 978-7-121-12357-3

I. ①无… II. ①杜… III. ①无线电通信:数据通信—高等学校—教材 IV. ①TN919.72

中国版本图书馆CIP数据核字(2010)第227352号

责任编辑:张来盛(zhangls@phei.com.cn)

印 刷:

装 订: 北京季蜂印刷有限公司

出版发行:电子工业出版社

北京市海淀区万寿路173信箱 邮编 100036

开 本: 787×1092 1/16 印张: 19.75 字数: 500千字

印 次: 2011年1月第1次印刷

印 数: 3000册 定价: 35.00元

凡所购买电子工业出版社图书有缺损问题,请向购买书店调换。若书店售缺,请与本社发行部联系,联系及邮购电话:(010)88254888。

质量投诉请发邮件至 zltts@phei.com.cn, 盗版侵权举报请发邮件至 dbqq@phei.com.cn。

服务热线:(010)88258888。

# 前 言

在过去的几十年间,无线数据通信按照自己的规则蓬勃发展。从蜂窝电话网到无线接入 Internet 和无线家庭网络等,无线数据通信给我们的生活带来了深刻影响,在经过呈指数增长后,今天的无线数据通信产业已经成为世界上最大的产业之一。

信息时代的到来使我们越来越依赖于无线通信,虽然便携式电话系统和个人通信系统使无线通信得到了广泛的应用,但最依赖无线通信的无疑还是军事领域,军事通信技术的发展始终走在民用技术的前面,并带动和促进民用通信。

当今的社会是一个高度信息化的社会。在信息社会中,信息的存储与处理离不开计算机,而信息的传递则离不开通信与计算机网络。无线数据通信正是建立在计算机网络基础之上,为实现移动终端与移动终端、移动终端与固定终端之间的数据信息传递而产生的通信方式。

无线数据通信与网络正在改变着我们的商务活动和生活方式,如电子商务、电子邮件、语音信箱、远程教育、视频聊天等,都是无线数据通信与网络技术的应用实例。无线数据通信与网络的发展目标是能够在世界的任何地点传输和交换诸如文本、音频和图像之类的数据。“5W”(Who、Whenever、Wherever、Whomever、Whatever)是无线数据通信发展的最终目标,是人们无拘束地获取和交互信息的一种内在需求,现代通信技术的发展事实上就是围绕“5W”这一目标逐步向前推进的过程。

无线和数据的融合,就是个人化的通信模式和宽带的通信能力以及丰富的通信内容的融合,是通信迈向“5W”发展的必然途径,也是当前通信发展朝着宽带化、智能化和个人化发展的主要潮流和必然趋势。

本书可作为现代无线数据通信技术参考书,也可作为一本现代军事战术通信技术的参考书,适合通信及相关专业的科研人员、教学人员、研究生、高年级本科生以及科研管理人员、装备检验人员、通信技术干部和指挥人员阅读使用。

全书共分 14 章。第 1 章绪论,介绍数据通信与网络的基本概念,包括数据通信系统的组成与特点,网络的拓扑结构,协议与标准,数据传输模式,多路复用,数据通信系统的性能指标等,为学好后述各章做好准备。

第 2 章无线通信信道,首先介绍无线信道的概念与分类,电波传播方式与无线信道特点,广义信道的数学模型,然后分析恒参信道、变参信道及其对所传信号的影响。最后说明信道带宽、系统带宽与信号带宽以及信道容量。总之,任何一个通信系统与网络,从大的方面均可视为由发送端、信道和接收端三大部分组成。因此,无线信道是无线数据通信系统与网络必不可少的组成部分,无线信道特性的好坏直接影响到系统与网络的总特性。

第 3 章信源与信源压缩编码技术,现代通信系统中信源可以分为三类:计算机产生的数据流、语音产生的数据流和图像产生的数据流,而数据信号的带宽要大于模拟信号的带宽,这将占用较大的信道资源。为了提高信道(特别是无线信道)的利用率和传输效率,在有限的带宽资源条件下,进行信源压缩编码是最有效的方法,信源压缩可以显著地减少传输所需的带宽,也就可以降低费用和发送时间。本章重点讨论了语音波形压缩编码技术、语音参量压缩编码技术、图像压缩编码技术与数据信号的压缩编码技术,并进一步介绍了语音、图像及数据压缩编码标准。

第4章数据基带传输系统与技术，数字和数据基带传输系统类似，所以本章不区别数字和数据的概念。在数据基带传输系统中，其传输对象通常是二元数字信息，它可能来自计算机、电传打字机或其他数字设备的各种数据代码，也可能来自数字电话终端的 PCM 脉冲编码信号等。由于未经调制的电脉冲数字信号所占据的频带通常从直流或低频开始，我们常称为基带信号。在某些有线信道中，特别是传输距离不太远的情况下，基带信号可以直接传送，我们称之为数据信号的基带传输。在另外一些信道，特别是无线信道和光信道中，数据基带信号则必须经过调制，将信号频谱搬移到高频处才能在信道中传输，我们把这种传输称为数据信号的调制传输（或载波传输、频带传输）。如果把无线调制与解调过程看作是广义信道的一部分，则任何数据传输系统均可等效为基带传输系统，因此分析掌握数据信号的基带传输原理是十分重要的。

第5章高性能数据调制技术，信号调制即对信号的调和整治，使信号匹配信道特性以适合信道传输。调制可实现频率分配、多路复用、减少噪声干扰以及方便辐射等。在信号调制系统中，通常把待传输消息对应的基带信号称为调制信号，调制系统输出的信号称为已调制信号，简称已调信号，用于承载基带信号传输的光波或电波称为载波。解调是调制的逆过程，就是从已调信号中提取消息信号的过程，这样可方便接收者（信宿）处理或解释。本章主要讨论最小频移键控（MSK）、正交振幅调制（QAM）、正交频分复用（OFDM）等高性能调制技术。

第6章抗干扰通信与通信干扰技术，无线抗干扰通信与通信干扰是一项涉及基础理论、技术体制、关键技术、性能评估、组织运用和战场管理控制等多方面的复杂系统工程，也是一门随着信息作战需求和通信电子进攻的挑战而发展起来的新兴工程学科，是当今武器装备发展的前沿研究领域。从哲学上讲，既没有无坚不摧的矛，也没有牢不可破的盾，抗干扰通信与通信干扰作为一对“矛”与“盾”的双方，是现代国防建设永恒的课题，通信抗干扰能力（尤其是复杂电磁环境下的通信抗干扰能力）是军队战斗力的重要标志。本章主要讨论抗干扰通信与通信干扰的基本理论、无线抗干扰通信技术、通信干扰技术。

第7章无线信道接入方法和多址技术，无线通信系统是以信道来区分对象的，一个信道只能容纳一个用户进行通话，许多同时通话的用户互相以信道来区分，这就是多址。无线通信是一个多信道同时工作的系统，具有广播和大面积覆盖的特点。在电波覆盖区内，如何建立用户之间的无线信道的连接是多址接入方式的问题，解决多址接入问题的方法即为多址技术。多址接入技术在无线通信中占有重要的地位，它关系到系统的构成、容量、频谱和信道利用率以及系统复杂性。本章主要讨论无线多址理论基础、频分多址技术、时分多址技术、码分多址技术、扩频及混合多址技术、正交频分复用（OFDM）多址接入、空间与极化多址方式。

第8章超宽带（UWB）通信技术，超宽带（UWB: Ultra-Wide Band）技术是利用超宽频带的电波进行高速无线通信的技术。从时域上讲，超宽带系统有别于传统的通信系统。一般的通信系统是通过发送射频载波进行信号调制，而 UWB 是利用起、落点的时域脉冲（几十 ns）直接实现调制。超宽带的传输把调制信息过程放在一个非常宽的频带上进行，而且以这一过程中所持续的时间来决定带宽所占据的频率范围。本章主要讨论超宽带基本概念、超宽带信号的基本调制方式、UWB 关键技术介绍、UWB 应用介绍。

第9章差错控制与信道编码技术，信道编码，又称差错控制编码、可靠性编码、抗干扰编码或纠错码，它是提高数字信号传输可靠性的有效方法之一。在无线数据传输过程中，引起数字传输差错的根本原因在于信道内存在的噪声，以及信道传输特性不理想所造成的码间串扰，为了提高数字传输系统的可靠性，降低信息传输的差错率，我们可以利用均衡技术消

除码间串扰,利用增大发射功率、降低接收设备本身的噪声、选择好的调制制度和解调方法、加强天线的方向性等措施,提高数字传输系统的抗噪性能,但上述措施受到各种条件制约,只能将传输差错减小到一定程度。若希望进一步提高数字传输系统的可靠性,就可以采用差错控制编码,对可能或已经出现的差错进行控制。本章主要讨论差错控制编码的基本概念、分组码、卷积码、TCM 网格编码调制、Turbo 码。

第 11 章无线数据通信协议,数据通信协议是网络内使用的“语言”,用来协调网络的运行,以达到互通、互控和互换的目的,通信的双方要共同遵守这些约定。在现代通信网中,一般用户通过应用软件使用数据信息,应用软件使用网络进行通信时并不直接同网络硬件打交道,而是同给定的协议规则打交道,通信网络协议是通信网络中不可缺少的重要组成部分。在无线数据通信网中,信道特性的非对称性、长时延、数据包丢失及比特差错的严重性等,使其有别于有线通信环境,其协议或规程不能直接照搬,因此编写本章内容。本章主要讨论开放系统互连参考模型、数据链路传输控制规程、无线数据通信网中的信道接入协议、无线局域网协议、无线宽带数据通信协议。

第 12 章无线数据通信组网技术,无线数据通信组网主要内容包括网络体系结构的设计和协议的设计,前者需要解决网络拓扑结构、网络控制管理、协议体系结构顶层设计问题,后者主要针对所设计的网络体系结构进行有效拓扑控制算法、管理协议、通信协议等底层技术的实现。本章主要讨论无线数据网络体系结构、蜂窝网络拓扑、移动 Ad Hoc 网络技术、无线 MESH 网络技术。

第 13 章战术数据链系统,战术数据链是连接数字化战场上的指挥中心、作战部队、武器平台的一种信息处理、交换和分发系统。“数据链”一词,最早起源于 20 世纪 50 年代末,当时即为北大西洋公约组织(NATO)及美国海军所使用。它主要是采用无线网络通信技术和应用协议,实现机载、陆基、舰载的指挥和控制时的战术数据信息交换。本章主要讨论数据链系统组成与数据传输流程、数据链组网协议、数据链的时隙分配。

第 14 章自由空间激光通信技术,自由空间激光通信又称无线光通信或自由空间光通信(Free Space Optical Communication, FSO),是光纤通信和无线通信相结合的产物,它是指以  $10^5$  GHz 以上的激光作为信息载体,不使用光纤等有线信道的传输介质,在宇宙或大气中传递信息的一种通信技术。本章主要讨论自由空间光通信系统基本组成、最新国内外研发进展与应用情况等。

本书是编者多年从事无线通信科研与教学的资料、经验和体会。为了适应教学需求,本书在编写过程中力求循序渐进,尽量保持叙述内容的完整性、实用性。

全书由杜思深主编,甘忠辉任副主编,西安电子科技大学葛建华教授主审,郭建新、薛凤凤、林宝勤、蒋磊等参编,吴静、李战军、姚昭琼、夏小梅、符鸿峰、杜菁、强小娟、夏智高等制作了插图和电子课件。在编写过程中,承蒙黄国策、达新宇、陈述新、庞宝茂、吕娜、刘剑等专家提供了大量宝贵的意见与建议,并参考和引用了许多参考资料。在此,向为本书的编写、出版工作作出贡献的所有人员深表感谢。

最后需要说明的是,由于本书涉及的内容广泛,编者水平有限,加之时间仓促,书中难免存在错误和不足,恳请读者批评指正。

本书有配套的电子课件,需要的读者可与作者联系,联系方式: du2877@163.com。

# 目 录

|                         |    |
|-------------------------|----|
| 第 1 章 绪论                | 1  |
| 1.1 无线数据通信基础            | 1  |
| 1.1.1 无线数据通信的发展         | 1  |
| 1.1.2 消息、信息与信息量         | 2  |
| 1.1.3 模拟、数字与数据          | 4  |
| 1.1.4 电路、信号、系统          | 5  |
| 1.1.5 无线数据通信系统与网络       | 6  |
| 1.2 无线数据通信方式与类型         | 10 |
| 1.2.1 无线数据通信与计算机通信、数字通信 | 10 |
| 1.2.2 无线数据通信的工作方式       | 11 |
| 1.2.3 无线数据通信的信号传输方式     | 11 |
| 1.2.4 无线数据通信的过程与信息流的方式  | 14 |
| 1.2.5 数据通信的业务类型         | 16 |
| 1.3 无线数据通信的主要性能指标       | 17 |
| 1.4 无线数据通信研究的主要内容与发展目标  | 19 |
| 思考与练习                   | 21 |
| 第 2 章 无线通信信道            | 22 |
| 2.1 无线信道的概念与分类          | 22 |
| 2.2 电波传播方式与无线信道特点       | 23 |
| 2.2.1 无线电波的传播方式         | 23 |
| 2.2.2 近地空间信道的特点         | 25 |
| 2.2.3 对流层散射信道的特点        | 27 |
| 2.2.4 电离层折射信道的特点        | 27 |
| 2.2.5 卫星通信信道的特点         | 29 |
| 2.3 广义信道的数学模型           | 30 |
| 2.3.1 狭义信道与广义信道         | 30 |
| 2.3.2 调制信道的数学模型         | 32 |
| 2.3.3 数字编码信道的数学模型       | 34 |
| 2.4 恒参信道及其对所传信号的影响      | 34 |
| 2.4.1 线性畸变与改善措施         | 34 |
| 2.4.2 非线性畸变和其他影响        | 36 |
| 2.4.3 噪声干扰              | 36 |
| 2.5 变参信道及其对所传信号的影响      | 36 |
| 2.5.1 多径传播、频率弥散和选择性衰落   | 37 |
| 2.5.2 多普勒频移与多普勒衰落       | 39 |

|       |                         |    |
|-------|-------------------------|----|
| 2.5.3 | 阴影效应与衰落                 | 39 |
| 2.5.4 | 变参信道特性的改善               | 39 |
| 2.6   | 信道带宽、系统带宽与信号带宽          | 41 |
| 2.6.1 | 带宽的概念与种类                | 41 |
| 2.6.2 | 带宽的几种计算方法               | 42 |
| 2.7   | 信道容量                    | 45 |
| 2.7.1 | 香农信道容量                  | 45 |
| 2.7.2 | 奈奎斯特信道容量                | 46 |
|       | 思考与练习                   | 46 |
| 第3章   | 信源与信源压缩编码技术             | 47 |
| 3.1   | 引言                      | 47 |
| 3.2   | 语音波形压缩编码技术              | 47 |
| 3.2.1 | 抽样定理及实际抽样的方法            | 48 |
| 3.2.2 | 量化及量化信号与噪声功率比           | 49 |
| 3.2.3 | 编码方案及比较                 | 53 |
| 3.2.4 | 脉冲编码调制 (PCM)            | 55 |
| 3.2.5 | 自适应差值脉冲编码调制 (ADPCM)     | 57 |
| 3.2.6 | 子带编码 (SBC)              | 59 |
| 3.3   | 语音参量压缩编码技术              | 61 |
| 3.3.1 | 语音产生的过程及数学模型            | 61 |
| 3.3.2 | 参量编码与声码器简介              | 63 |
| 3.3.3 | 线性预测声码器的原理              | 64 |
| 3.4   | 混合压缩编码技术                | 67 |
| 3.4.1 | 多脉冲激励线性预测声码器 (MPLPC)    | 68 |
| 3.4.2 | 规则脉冲激励线性预测声码器 (PRE-LPC) | 68 |
| 3.4.3 | 码激励线性预测声码器 (CELP)       | 69 |
| 3.5   | 语音及音频压缩编码标准             | 71 |
| 3.5.1 | 语音压缩编码标准                | 71 |
| 3.5.2 | 音频压缩编码标准                | 72 |
| 3.5.3 | 音频压缩文件的常见格式             | 75 |
| 3.6   | 图像的抽样量化与压缩编码            | 76 |
| 3.6.1 | 概述 (图像信号的特征)            | 76 |
| 3.6.2 | 图像信号的抽样与量化              | 78 |
| 3.6.3 | 图像压缩编码基本方法              | 81 |
| 3.7   | 图像压缩编码标准                | 85 |
| 3.7.1 | 静止图像压缩编码标准              | 85 |
| 3.7.2 | 活动图像压缩编码标准              | 86 |
| 3.8   | 数据信号的压缩编码               | 88 |
| 3.8.1 | 数据信号的压缩编码的特点            | 88 |

|                                 |            |
|---------------------------------|------------|
| 3.8.2 数据信号的压缩编码方法 .....         | 89         |
| 思考与练习 .....                     | 93         |
| <b>第 4 章 数据基带传输系统与技术</b> .....  | <b>94</b>  |
| 4.1 数据基带信号编码技术 .....            | 94         |
| 4.1.1 数据基带信号的编码原则 .....         | 94         |
| 4.1.2 二元码 .....                 | 95         |
| 4.1.3 三元码 .....                 | 98         |
| 4.1.4 多元码 .....                 | 100        |
| 4.2 无码间串扰的基带系统特性 .....          | 101        |
| 4.2.1 基带系统组成及各点波形 .....         | 101        |
| 4.2.2 基带传输特性与冲激响应 .....         | 102        |
| 4.2.3 系统无码间串扰的条件 .....          | 103        |
| 4.3 基带传输系统的性能改进技术介绍 .....       | 104        |
| 4.3.1 部分响应技术 .....              | 104        |
| 4.3.2 时域均衡技术 .....              | 106        |
| 4.4 眼图 .....                    | 111        |
| 思考与练习 .....                     | 113        |
| <b>第 5 章 高性能数据调制技术</b> .....    | <b>115</b> |
| 5.1 引言 .....                    | 115        |
| 5.1.1 调制技术的主要性能指标 .....         | 115        |
| 5.1.2 无线调制技术的发展与分类 .....        | 116        |
| 5.2 最小频移键控 (MSK) .....          | 118        |
| 5.2.1 最小频移键控的基本原理 .....         | 118        |
| 5.2.2 高斯滤波最小频移键控 (GMSK) .....   | 123        |
| 5.3 正交振幅调制 (QAM) .....          | 124        |
| 5.3.1 正交振幅调制的基本原理 .....         | 124        |
| 5.3.2 可变速率正交振幅调制 (VR-QAM) ..... | 126        |
| 5.4 正交频分复用 (OFDM) .....         | 128        |
| 5.4.1 OFDM 基本原理 .....           | 128        |
| 5.4.2 用离散傅里叶变换实现 OFDM .....     | 130        |
| 5.4.3 OFDM 的优缺点 .....           | 132        |
| 思考与练习 .....                     | 132        |
| <b>第 6 章 抗干扰通信与通信干扰技术</b> ..... | <b>133</b> |
| 6.1 引言 .....                    | 133        |
| 6.1.1 通信对抗基本概念 .....            | 133        |
| 6.1.2 抗干扰通信的基本概念 .....          | 134        |
| 6.1.3 通信干扰基本概念 .....            | 135        |
| 6.2 抗干扰通信技术 .....               | 135        |

|                                   |            |
|-----------------------------------|------------|
| 6.2.1 抗干扰通信的理论基础 .....            | 135        |
| 6.2.2 扩谱抗干扰通信技术 .....             | 138        |
| 6.3 通信干扰技术 .....                  | 146        |
| 6.3.1 通信干扰技术的分类 .....             | 146        |
| 6.3.2 通信干扰的理论基础 .....             | 148        |
| 6.3.3 对常规通信信号的干扰 .....            | 151        |
| 6.3.4 对扩谱抗干扰信号的干扰 .....           | 155        |
| 思考与练习 .....                       | 159        |
| <b>第 7 章 无线信道接入方法和多址技术</b> .....  | <b>160</b> |
| 7.1 无线多址理论基础 .....                | 160        |
| 7.1.1 无线双工通信方式 .....              | 160        |
| 7.1.2 无线信号分割原理 .....              | 161        |
| 7.2 频分多址技术 .....                  | 162        |
| 7.2.1 FDMA 基本原理 .....             | 162        |
| 7.2.2 FDMA 系统中的干扰问题 .....         | 162        |
| 7.2.3 FDMA 系统的特点 .....            | 163        |
| 7.3 时分多址技术 .....                  | 164        |
| 7.3.1 TDMA 基本原理 .....             | 164        |
| 7.3.2 TDMA 的帧结构 .....             | 165        |
| 7.3.3 TDMA 系统的特点 .....            | 165        |
| 7.3.4 TDMA 系统的帧效率与系统容量 .....      | 166        |
| 7.4 码分多址技术 .....                  | 166        |
| 7.4.1 CDMA 基本原理 .....             | 166        |
| 7.4.2 CDMA 系统的特点 .....            | 168        |
| 7.4.3 CDMA 系统的两个问题 .....          | 169        |
| 7.5 扩频及混合多址技术 .....               | 170        |
| 7.5.1 跳频多址 .....                  | 170        |
| 7.5.2 混合扩频多址 .....                | 171        |
| 7.6 正交频分复用多址接入 .....              | 174        |
| 7.6.1 OFDM-TDMA .....             | 174        |
| 7.6.2 OFDM-CDMA .....             | 174        |
| 7.6.3 OFDM-FDMA .....             | 175        |
| 7.7 空间与极化多址方式 .....               | 175        |
| 7.7.1 空分多址 (SDMA) .....           | 175        |
| 7.7.2 极化区分多址 (PDMA) .....         | 176        |
| 思考与练习 .....                       | 177        |
| <b>第 8 章 超宽带 (UWB) 通信技术</b> ..... | <b>179</b> |
| 8.1 超宽带基本概念 .....                 | 179        |
| 8.1.1 超宽带定义及规范 .....              | 179        |



|        |                            |     |
|--------|----------------------------|-----|
| 8.1.2  | 超宽带通信与现有通信的比较              | 180 |
| 8.1.3  | 超宽带通信的特点                   | 181 |
| 8.2    | 超宽带信号的基本调制方式               | 183 |
| 8.2.1  | 基于跳时体制的脉冲位置调制 (UWB TH-PPM) | 183 |
| 8.2.2  | 基于跳时体制的脉冲幅度调制 (UWB TH-PAM) | 185 |
| 8.2.3  | 基于脉冲的直接序列调制 (DS-UWB)       | 186 |
| 8.3    | UWB 关键技术介绍                 | 187 |
| 8.3.1  | 脉冲信号的产生技术                  | 187 |
| 8.3.2  | UWB 的调制技术                  | 187 |
| 8.3.3  | UWB 多址技术                   | 188 |
| 8.4    | UWB 应用介绍                   | 189 |
|        | 思考与练习                      | 191 |
| 第 9 章  | 差错控制与信道编码技术                | 192 |
| 9.1    | 差错控制编码的基本概念                | 192 |
| 9.1.1  | 差错控制编码与信源编码的区别             | 192 |
| 9.1.2  | 错误分类与错误图样                  | 192 |
| 9.1.3  | 差错控制方式                     | 193 |
| 9.1.4  | 最小码距与检 (纠) 错能力             | 195 |
| 9.2    | 分组码                        | 197 |
| 9.2.1  | 线性分组码                      | 197 |
| 9.2.2  | 循环码                        | 200 |
| 9.3    | 卷积码                        | 203 |
| 9.3.1  | 卷积码编码器及表述方法                | 203 |
| 9.3.2  | 卷积码的译码                     | 208 |
| 9.4    | 网格编码调制                     | 215 |
| 9.4.1  | 引言                         | 215 |
| 9.4.2  | 集合分割原理                     | 216 |
| 9.4.3  | TCM 码网格图的构造                | 218 |
| 9.4.4  | TCM 码的卷积码表示                | 219 |
| 9.5    | Turbo 码                    | 220 |
| 9.5.1  | 编码结构和原理                    | 220 |
| 9.5.2  | 译码器结构和原理                   | 221 |
|        | 思考与练习                      | 222 |
| 第 10 章 | 通信加密解密与密码分析                | 223 |
| 10.1   | 密码学基本概念                    | 223 |
| 10.1.1 | 密码学发展历史                    | 223 |
| 10.1.2 | 密码体制模型                     | 225 |
| 10.1.3 | 密码体制的分类                    | 226 |
| 10.1.4 | Shannon 的保密通信理论            | 228 |

|        |                                 |     |
|--------|---------------------------------|-----|
| 10.2   | 解密与密码分析.....                    | 231 |
| 10.2.1 | 解密和密码分析.....                    | 231 |
| 10.2.2 | 密码分析的方法.....                    | 232 |
| 10.3   | 通信系统与网络保密技术.....                | 233 |
| 10.3.1 | 语音保密通信技术.....                   | 233 |
| 10.3.2 | 图像保密通信技术.....                   | 235 |
| 10.3.3 | 数据通信网络保密技术.....                 | 237 |
|        | 思考与练习.....                      | 240 |
| 第 11 章 | 无线数据通信协议.....                   | 241 |
| 11.1   | 数据通信协议的概念.....                  | 241 |
| 11.1.1 | 协议的概念及功能.....                   | 241 |
| 11.1.2 | 开放系统互连参考模型.....                 | 242 |
| 11.2   | 数据链路传输控制规程.....                 | 245 |
| 11.2.1 | 数据链路控制规程的功能与分类.....             | 245 |
| 11.2.2 | 面向字符型的传输控制规程.....               | 249 |
| 11.2.3 | 面向比特型的传输控制规程.....               | 251 |
| 11.3   | 无线数据通信网中的信道接入协议.....            | 258 |
| 11.3.1 | 单信道无线接入协议.....                  | 258 |
| 11.3.2 | 多信道接入协议.....                    | 259 |
| 11.3.3 | 无线信道的隐终端和暴露终端问题.....            | 261 |
| 11.4   | 无线局域网协议.....                    | 262 |
| 11.4.1 | 无线局域网解决方案与运作模式.....             | 262 |
| 11.4.2 | IEEE 802.11 标准中的物理层.....        | 263 |
| 11.4.3 | IEEE 802.11 标准中的 MAC 子层.....    | 263 |
| 11.4.4 | 载波侦听多址访问/碰撞回避机制.....            | 264 |
| 11.4.5 | RTS/CTS 协议.....                 | 265 |
| 11.4.6 | 随机的时间等待 IFS.....                | 267 |
| 11.5   | 无线宽带数据通信协议.....                 | 268 |
| 11.5.1 | IEEE 802.16/WiMAX 标准化进程及发展..... | 268 |
| 11.5.2 | IEEE 802.16e 协议结构.....          | 269 |
|        | 思考与练习.....                      | 273 |
| 第 12 章 | 无线数据通信组网技术介绍.....               | 274 |
| 12.1   | 无线数据组网的发展.....                  | 274 |
| 12.2   | 无线数据网络体系结构.....                 | 275 |
| 12.3   | 蜂窝网络拓扑.....                     | 276 |
| 12.4   | 移动 Ad Hoc 网络技术.....             | 277 |
| 12.5   | 无线 Mesh 网络技术.....               | 279 |
|        | 思考与练习.....                      | 282 |

第 13 章 战术数据链系统与技术..... 283

13.1 战术数据链基本概念 ..... 283

13.1.1 何谓战术数据链..... 283

13.1.2 战术数据链功能及特点 ..... 284

13.1.3 战术数据链的分类..... 286

13.2 战术数据链系统组成与数据传输流程 ..... 286

13.2.1 战术数据链系统组成..... 286

13.2.2 数据传输流程..... 287

13.3 战术数据链组网协议 ..... 289

13.3.1 有中心节点的组网协议 ..... 289

13.3.2 无中心节点的组网协议 ..... 290

13.4 战术数据链的时隙分配 ..... 291

13.4.1 链路终端的入网过程..... 291

13.4.2 时隙及时隙分配..... 292

13.4.3 时隙传输结构..... 293

13.4.4 数据封装结构..... 293

思考与练习 ..... 294

第 14 章 自由空间激光通信技术..... 295

14.1 自由空间光通信系统的基本组成..... 295

14.2 FSO 的技术特点与优势..... 297

14.3 最新国内外研发进展与应用情况..... 298

14.3.1 国外研发的最新进展与应用情况..... 298

14.3.2 国内研发的进展与应用情况 ..... 299

14.4 存在的问题和解决思路 ..... 300

思考与练习 ..... 301

参考文献 ..... 302

# 第 1 章 绪 论

随着科学技术的发展，从传统的电报、电话到今天的卫星通信、移动通信与计算机互联网，通信已经和我们的日常生活紧密地结合在一起了，并已深入到了社会的每一个角落，进入了千家万户，每个人都从通信中得到了服务和方便。特别是计算机技术和通信的结合，使得通信越来越受到广大用户的欢迎。现在，“信息”在人类社会的“物质、能源、信息”三大基础要素中的地位已上升到一个崭新的高度，并成为推动人类社会文明、进步与发展的巨大动力。

通过本章的学习，建立起无线数据通信的基本概念，了解无线数据通信有关的标准，以及世界主要的标准化组织等，为后续章节的学习打下基础。

## 1.1 无线数据通信基础

### 1.1.1 无线数据通信的发展

按照人类通信交流方式与技术的不同，可以把通信的历史划分为 6 个阶段：

- 第一阶段的通信方式是语言，人们通过人力、马力和烽火台等原始通信手段传递信息。
- 第二阶段从发明文字和邮政通信开始。
- 第三阶段以发明印刷术为标志。
- 第四阶段从 1838 年发明莫尔斯电报开始，人们进入了电气通信的初级阶段。由于此时的通信是以电信号来远距离传递消息，所以称为电信（Telecommunication）。
- 第五阶段是以 1948 年香农（Shannon）提出的信息论开始的近代通信阶段。
- 第六阶段是以 20 世纪 80 年代出现的光纤通信系统（Optical Fiber Communication）为代表以及以移动通信和综合业务数字网的迅速崛起为标志的现代通信阶段。

数据通信并不是现在才出现，也许可以追溯到第一阶段或“烽火戏诸侯”的时代。夜晚点火或白天点狼烟表示有敌人入侵，可称为“1”状态；无火或没有狼烟表示平安无事，称为“0”状态。海军的旗语是用两只手的握旗姿势表示 A~Z 等符号，这也可算是多进制的数据通信了。

无线电数据通信，也许可从 1838 年莫尔斯发明的电报（莫尔斯电报，以几个阿拉伯数字表示一个字符）算起，但作为真正意义的无线数据通信要从 20 世纪 80 年代无线数据接入和移动通信算起，即无线数据通信伴随着通信技术的进步而共同发展。

纵观通信技术特别是无线电数据通信的发展历程，可以看出，经历了点到点的通信，再到多点之间的信息传输与交换，最后进入网络时代的发展过程。

现代通信一般为各种网络用户之间的通信，与传统点到点直接通信的概念不同，通常我们把从一个地方向另一个地方进行消息（信息）的迅速、准确、有效、可靠传送与（或）交换，称为通信。这里的“迅速”指通信的及时性，“准确”指通信目标的准确性，“有效”指

通信的传输效率，“可靠”指通信的差错率。传送或交换信息所需的一切技术设备的总和称为通信系统。

### 1.1.2 消息、信息与信息量

无线数据通信的任务，就是将消息以信号形式传送给接收者，并使接收者获得最多的信息。那么什么是消息、信号与信息？它们有什么不同？

#### 1. 消息

消息 (Message) 是通信系统中传输的具体对象，是信号的具体内容，是表达信息的形式。消息的形式是多样的，包括符号、文字、语音、数据、图像、视频等。同一种信息的内容可用不同形式的消息来表达。例如，天气预报可用文字消息表达，也可用语音消息来表达。

#### 2. 信息

信息 (Information) 指的是消息所包含的内容。例如，每天的天气预报是一种消息，预报中告知某日某时的真实天气情况如何，就是该消息所包含的“信息”。再如，在一次传真通信中，甲方给乙方传真了一份合同，乙方收到的合同上除了文字外还有花纹和图案等，但对乙方而言，只有文字有意义；也就是说，对乙方而言，信息是文字，消息是整个合同。

消息有可能包含甚为丰富的信息，但也可能信息甚少，若一种消息并未给人们带来新的知识，那么这种消息所包含的信息实际等于零。

所以，信息是消息中对接收者有意义的部分，也就是消息中所包含的收信者原来不知道而有待知道的内容，即消息的内涵，消息是外壳。消息可以有各种各样的形式，但消息的内容可统一用信息来表述。

信息的传送一般都不是直接的，它必须借助于一定形式的信号（光信号、电信号等），才能传输和进行各种处理。

#### 3. 信息量及平均信息量“熵”

在一切有意义的通信中，虽然消息的传递意味着信息的传递，但对接收者而言，某些消息比另外一些消息的传递具有更多的信息。例如，甲方告诉乙方一件经常可能发生的事情“明天中午 12 时正常开饭”，比起告诉乙方一件极不可能发生的事情“明天 12 时有地震”来说，前一消息所包含的信息显然要比后者少些。这是因为对乙方（接收者）来说，前一事情很可能（或必然）发生，不足为奇，而后一个事情却极难发生，听后会使人惊奇。这表明消息确实有量值的意义，而且可以看出，对接收者来说，事件越不可能发生，越会使人感到意外和惊奇，则信息量就越大。

正如已经指出的，消息是多种多样的，因此，度量消息中所含的信息量值，必须能够用来估计任何消息的信息量，且与消息种类无关。另外，消息中所含信息的多少也应和消息的重要程度无关。

由概率论可知，事件的不确定程度，可用事件出现的概率来描述，事件出现（发生）的可能性越小，则概率越小；反之，概率越大。

基于这种认识，我们得到：消息中的信息量与消息发生的概率紧密相关。消息出现的概

率越小，则消息中包含的信息量就越大，且概率为零时（不可能发生事件）信息量为无穷大，概率为 1 时（必然事件）信息量为 0。

综上所述，消息  $x$  中所含信息量  $I$  是消息出现的概率  $P(x)$  的函数，即：

$$I(x) = \log_a \frac{1}{P(x)} = -\log_a P(x) \tag{1-1}$$

信息量  $I$  的单位与对数的底数有关：

- $a=2$ ，单位为比特（bit，简写为 b）；
- $a=e$ ，单位为奈特（nat，简写为 n）；
- $a=10$ ，单位为笛特（Det）或称为十进制单位。

通常使用的单位为 bit。下面我们以举例形式说明简单信息量的计算。

**例 1-1** 试计算二进制符号等概率和  $M$  进制等概率时每个符号的信息量。

**解：**二进制等概率时，有  $P(1)=P(0)=1/2$ ，则

$$I(1) = I(0) = -\log_2 \frac{1}{2} = 1 \text{ bit}$$

$M$  进制等概率时，有  $P(1)=P(2)=\cdots=P(M)=1/M$ ，则

$$I(1) = I(2) = \cdots = I(M) = -\log_2 \frac{1}{M} = \begin{cases} 2 \text{ bit}, & M = 4 \\ 3 \text{ bit}, & M = 8 \\ 4 \text{ bit}, & M = 16 \end{cases}$$

若每个符号出现的概率不等，则其信息量不同。这时需用平均信息量“熵” $H(x)$  的概念。由于信源中的各个符号是随机出现的，所以平均信息量只能用统计平均的方法得到，则  $H(x)$  等于各个符号的信息量乘以各自出现的概率再相加。“熵”的单位为 bit/每符。

$$H(x) = P(x_1) \log_2 \frac{1}{P(x_1)} + P(x_2) \log_2 \frac{1}{P(x_2)} + \cdots + P(x_n) \log_2 \frac{1}{P(x_n)} \tag{1-2}$$

**例 1-2** 设由 4 个符号 A、B、C、D 组成的信息源，相应概率为

|               |               |               |               |
|---------------|---------------|---------------|---------------|
| A             | B             | C             | D             |
| $\frac{1}{4}$ | $\frac{1}{8}$ | $\frac{1}{8}$ | $\frac{1}{2}$ |

其中  $P(A)+P(B)+P(C)+P(D)=1$ ，试求信源的平均信息量  $H(x)$ 。

**解：**A 符号出现包含的信息量为  $I(A)=-\log_2(1/4)=2 \text{ bit}$

B 符号出现包含的信息量为  $I(B)=-\log_2(1/8)=3 \text{ bit}$

C 符号出现包含的信息量为  $I(C)=-\log_2(1/8)=3 \text{ bit}$

D 符号出现包含的信息量为  $I(D)=-\log_2(1/2)=1 \text{ bit}$

则 4 个符号信息源的平均信息量为

$$\begin{aligned} H(x) &= P(A)I(A) + P(B)I(B) + P(C)I(C) + P(D)I(D) \\ &= \frac{1}{4} \times 2 + \frac{1}{8} \times 3 + \frac{1}{8} \times 3 + \frac{1}{2} \times 1 \\ &= 1.75 \text{ bit/每符} \end{aligned}$$

从例 1-1、例 1-2 可以看出，符号不等概率出现的平均信息量小于等概率出现的平均信息量，所以能保障等概率出现的符号传输，一定也能保障不等概率出现的符号传输。因此，在没有特殊说明时，书中所提的信息量都为平均信息量。

### 1.1.3 模拟、数字与数据

消息可分为模拟消息（Analog Message）、数字消息（Digital Message）和数据消息（Data Message）。

信息可分为模拟信息（Analog Information）、数字信息（Digital Information）和数据信息（Data Information），它们分别与模拟消息、数字消息和数据消息对应。那么什么是模拟、数字与数据？它们有什么不同？

#### 1. 模拟

模拟，即模仿、仿效。模拟消息是指消息的模仿与仿效，模拟消息的取值与原消息完全一致，有无穷多种可能，并与消息在时间上是否连续变化无关。例如，人嘴发出的声音，任何人在任意时刻的声音大小的取值有无穷多种可能，并在谈话停顿间隙，不出声，故在时间轴上非连续变化，但仍是典型的模拟消息。再如，人的体温在一定范围内有无穷多种可能取值，若每天测量 2 次，则体温读数在时间轴上是非连续变化，而我们仍然认为人体的温度是模拟量消息，这是因为，每次的测量结果有无穷多种可能。

#### 2. 数字

数字是表示数目的文字或符号。数字消息即用数目字表达的消息，或者说取值为有限数目字的消息。例如阿拉伯数字，只用 0、1、2、3、4、5、6、7、8、9 共 10 个符号；单路语音信号的二进制数字化消息，其特点是仅用有限位二进制数字编码表达语音，常见的有 8 位二进制数字编码。同样，数字消息与消息在时间上是否连续变化无关，仅与消息的表达取值的多少有关，只要能用有限位数目字表达的消息就可称为数字消息。

数字通信与模拟通信相比较主要有以下的优点：

- 抗干扰能力强，并在数字接力通信时可以消除噪声的积累；
- 可以利用数字纠错编码技术实现差错控制；
- 易于实现数字加密；
- 易于和大规模集成电路技术、计算机信号处理技术等现代技术相结合。

#### 3. 数据

数据，人们几乎每天都要接触到它，例如各种实验数据、各类统计报表等，它们是各种统计、计算、科学研究或技术设计等所依据的数值。尽管人们经常处理数值数据，但对数据还没有统一的严格定义。

通常意义上的“数据”是预先约定的、具有某种含义的任何一个数字或一个字母（符号）以及它们的组合构成的文字或数字记录的各种量值或内容等。

例如，我们常用的十进制阿拉伯数字位置计数法则，可以将三个数字“9”组合成数据“999”，读作“九百九十九”，而非“二十七”（ $9+9+9=27$ ）。虽然这里都是数字 9，但每个 9 的位置不同，代表的数据大小就不同。

再如，约定用正脉冲表示二进制数字“1”，表示电路接通，用负脉冲表示二进制数字“0”表示电路开断。这里，数字“1”和“0”就是数据。

现在，数据的来源、内容相当广泛，几乎涉及一切，最终以离散的数字信号表示的可被送到计算机中进行处理的各种信息，例如一份资料、一篇论文、一些图纸，甚至人的思维、语音及活动图像等信息编码都包括在内。因此，数据的概念逐渐从狭义过渡到广义的理解和应用。二进制编码的字母/数字符号、软件处理中的操作代码、控制代码、用户地址、程序数据或数据库信息都是数据。因此，数据是被处理、加工和存储的信息，也是消息的一种表达形式。

数据的表示形式是离散的数字，单个数字也可称为数据，即数字是数据的特例，但此数字要具有特殊的意义，这就是数据和数字的区别。

#### 1.1.4 电路、信号、系统

电路、信号、系统与网络的理论和分析方法，潜在的和实际的应用范围不断地在扩大着，几乎渗透到各个科学技术领域之中。那么，什么是电路（Circuit）？什么是信号（Signal）？什么是系统（System）？什么是网络（Network）？这是首先必须弄清楚的问题。

##### 1. 电路

通常，由电源、用电器、导线、电器元件等连接而成的电流通路，称为电路。

在电路系统中，由若干元件组成的用来使电信号按一定要求传输的电路或其中的部分，亦称为网络，当研究一般性的抽象规律时往往用网络一词，而讨论一些指定的具体问题时常称之为电路。

##### 2. 信号

信号这一术语不仅出现于科学技术领域之中，而且人们在日常生活之中每时每刻都与信号打交道，人们对信号并不陌生。

信号是用来携带消息的载体，表现为随时间变化的某种物理量。一般来说，用光线、电波、声音、动作、标志等传送的约定通信符号称为信号，用来指挥行动或指示目标，如口令、汽笛、红绿灯等。

一般将与信息相对应的随时间变化的电流或电压，也可以是电容的电荷，线圈的磁通以及空间的无线电磁波等称为电信号，简称信号。

信号是以数学的方法描述和表达实际通信系统中消息的形式，使之变成适合在信道上传输的信号。

信号特性可以从两个方面来描述：一是时间特性，亦称为时域特性，二是频率特性，亦称为频域特性。信号的时间特性与频率特性之间具有一定的相互对应关系。不同的时域特性将导致不同的频域特性。信号时间特性和频率特性不同，其形式也会不同。

依据信号在时间坐标上的取值特点不同，可将信号分为模拟信号（Analog Signal）、数字信号（Digital Signal）与数据信号（Data Signal）。

虽然数据的基本形式是离散的数字，但与数字信号的取值特点不完全一致，我们这里进行单独分出。其中为模拟信号、数字信号和数据信号，它们分别与模拟消息、数字消息和数据消息对应。



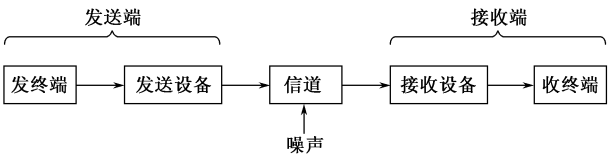
3. 系统

从一般意义上来说，所谓系统是指由若干个相互联系、相互作用的单元（事物）组合而成，为完成某个任务的特定功能整体。

通信系统的任务是完成消息的传递或交换。

构成系统的单元可小可大，可简可繁。如果将通信系统、控制系统、计算机系统与指挥系统共同组合而成一个繁杂的整体，可以构成一个综合系统。一只电阻和一只电容可以构成具有一定微分或积分功能的简单系统。

以点对点直接通信系统为例，要实现消息从一端向另一端的传递，不论通信系统是何种形式、传输何种的信号，都必须有 3 个部分：一是发送端，二是接收端，三是收，发两端之间的信道，其中发送端包括发终端与发送设备，接收端包括收终端与接收设备，如图 1-1 所示，这是通信系统的高度抽象概括的模型。



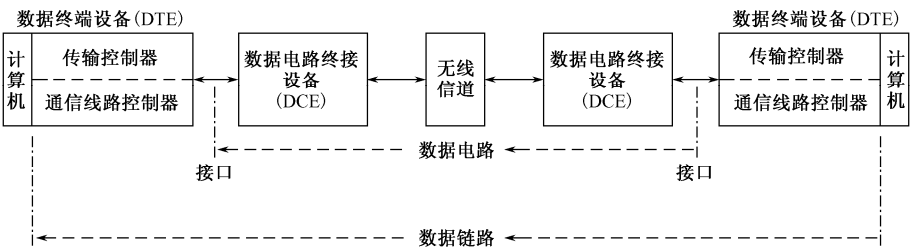
1.1.5 无线数据通信系统与网络

无线数据通信系统是指依据数据通信协议，利用数据传输技术，通过无线电数据链路将分布在远端的数据终端设备连接起来，实现数据信号无线传输、交换、存储和处理的系统。

无线数据通信中数据信号的传输、处理、调制解调电路等，与无线数字通信中数字信号传输、处理、调制解调电路等基本一样，无线数据通信并具有一切无线数字通信的一切优点，这并不能说无线数据通信等同于无线数字通信，因为数据通信要涉及更多的数据协议、规约等内容。

1. 无线数据通信系统基本组成

无线数据信号通信系统的具体组成可以各异，但其基本功能组成可以概括如图 1-2 所示，其各部分功能与作用如下。



1) 数据终端设备（Data Terminal Equipment，DTE）

DTE 一般由数据输入设备（产生数据的数据源）、数据输出设备（接收数据的数据宿）

和控制器（传输控制和通信线路控制）组成。

DTE 是一个总称，它主要包括计算机、打印机、电传机、智能仪表、传真机、数字电话机等，根据实际需要采用不同的设备。例如，在发送数据中，DTE 可以是键盘、鼠标和扫描仪等输入器；在接收数据中，它可以是显示器、打印机和传真机等。

DTE 的种类很多，按照使用场合可以分为通用数据终端和专用终端；按照性能可以分为简单终端和智能终端（如计算机等）。

DTE 在数据通信中的作用有点类似于电话与电报通信中的电话机和电传机，它把数据消息变成数据信号，并把这些数据信号输送到远端的数据终端（例计算机系统），同时，可以接收远端计算机系统的处理结果——数据，并将它变为人们能理解的数据消息。所以，DTE 相当于人和机器（计算机）之间的接口。

（1）传输控制器。传输控制器按照约定的数据通信控制规程，控制数据的传输过程。例如，收发之间的同步、传输错误的检测与纠正以及数据流串/并转换、数据缓冲、速度匹配等，以达到收发之间的协调、可靠地工作。

（2）通信线路控制器。通信线路控制器又称为前置处理机，控制与远端数据终端设备连接的全部通信信道（所有通信线路），接收远端 DTE 发来的数据信号，并向远端 DTE 发送数据信号。

通信控制器的主要功能，对远端 DTE 一侧来说，是差错控制、终端的接续控制、确认控制、传输顺序控制和切断等控制；对近端 DTE 一侧来说，其功能是将线路上传来的串行信号变成并行信号，或将计算机输出的并行信号变成串行信号。

## 2）数据电路终端设备（Data Circuit Terminal Equipment, DCE）

DCE 是 DTE 与传输信道的接口设备。

发方的 DCE 有两项功能：一是将终端送来原始数据信号变成更适合实际信道传输的信号形式，例如完成数据信号的变换、调制和纠错编码等；二是建立、保持和释放线路连接等功能。

收方的 DCE 则实现与发方相反的功能。

常用的电话线路调制解调器（Modem）就是一种数据电路终接设备，它是调制器和解调器的结合。发送时，调制器把数据信号转换成适合于电话线路上传输的模拟信号；接收时，模拟信号由解调器将它还原成数据信号，并送到 DTE。当数据信号在数字线路上传输时，DCE 的位置上不再需要 Modem，而改为数据服务单元（Data Service Unit, DSU）。DSU 的主要功能是信号码型与电平的转换、信道特性的均衡、收发时钟的形成与供给，以及线路接续控制等。

## 3）无线信道（Radio Channel）

无线信道是指传送信号的无线媒介或通路，可以是无线模拟信号通路，也可以是无线数字信号通路；可以是直达无线线路，也可以是无线中继、或通过交换设备沟通的无线通路。

## 4）数据电路（Data Circuit）/数据链路（Data Link）

数据电路由传输信道及其两端的数据电路终端设备（DCE）组成，不管是无线信道或有线信道，含有数据电路终端设备的广义信道，可以认为是数据信号直接传输的物理通路，简

称数据电路，可实现两个数据终端设备之间直接通信。这里数据电路仅指无线信道及其两端的数据电路终端设备组成的电路系统。

数据电路加上数据传输控制器的功能后就构成了数据链路，并将无线信道加上数据传输控制器的功能称为无线数据链路。对于数据终端设备来说，数据链路是数据终端的一个广义信道。

5) 接口（Interface）

接口是相关区域、实体、物质或阶段之间所形成共同界限的面。数据终端设备与数据电路终接设备存在接口，控制器与计算机之间也存在接口。

为了保证相同设备接口的一致性，接口标准规定了机械特性、电气特性、功能特性和规程特性等技术条件。

6) 协议组（族）

协议是用来管理和控制数据通信的一组（族）规则。它代表了通信设备之间的约定，如果没有协议，即使两个设备相连，也不能交流信息。

2. 无线数据通信网络的拓扑结构

系统与网络的差异，主要体现在观察事物的着眼点或处理问题的角度方面。系统问题侧重全局，而网络问题则侧重局部。例如，仅由电阻和电容组成的简单电路，在网络分析中，注意研究其各支路、回路的电压或电流；而从系统的观点来看，可以研究它如何构成具有微分或积分功能的运算器。

由于通信技术的发展，通信网络已形成了一个覆盖整个地球、具有各类网络和系统的非常复杂的体系。不同国家、不同地区、不同时期建立的各种网络汇聚在一起，人们尽最大的努力使这些网络融为一体，相互连通。例如，由分布在各地的许多模拟电话机、数字电话机（手机）、模拟信号通路、数字信号通路、市内数字交换设备、移动交换设备、多路复用设备、卫星通信信道、微波通信信道、光缆信道等所构成的网络，完成多点对点多的通信。

本书重点讨论无线数据通信网络，其拓扑结构主要有以下几种形式。

1) 直接连接

无线数据终端直接连接结构如图 1-3 所示，可取代有线连接，实现点对点通信，例如：红外线 IrDA。



图 1-3 无线数据终端直接连接结构示意图

2) 星状网络

无线数据终端星状网络连接结构如图 1-4 所示，它拥有一个中心路由和控制点，所有数

据都经过中心控制点，支持点对点、点对多点通信，如 WiFi、GSM、Bluetooth 等。

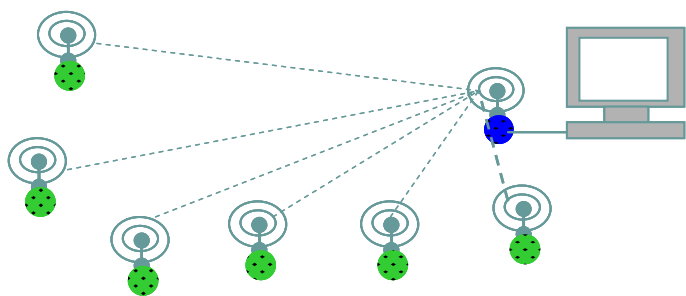


图 1-4 无线数据终端星状网络连接结构示意图

3) 树状网络

无线数据终端树状网络连接结构如图 1-5 所示，其特性和星状网络类似。

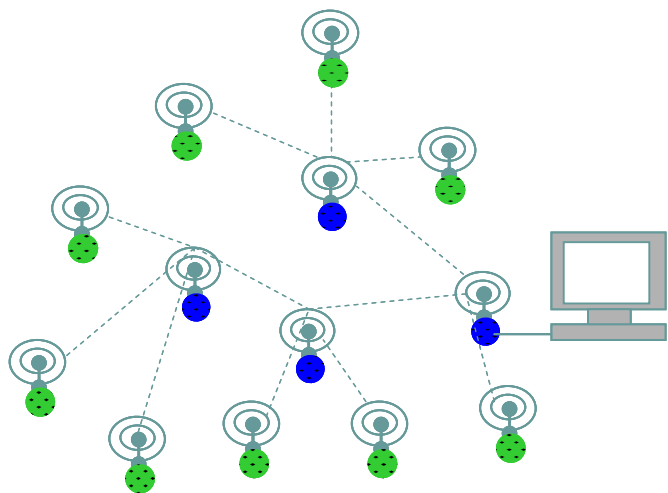


图 1-5 无线数据终端树状网络连接结构示意图

4) 网状网络

无线数据终端网状网络连接结构如图 1-6 所示，可实现多路径数据传输，自行配置和修复，如 ZigBee。

5) 蜂窝状网络

在蜂窝移动通信网中，为便于网络组织，将一个移动通信网分为若干个服务区，每个服务区又分为若干个移动交换中心区，每个移动交换中心区又分为若干个位置区，每个位置区由若干个基站小区组成。一个移动通信网由多少个服务区或多少个移动交换中心区组成，取决于移动通信网所覆盖地域的用户密度和地形地貌等。多个服务区的移动数据终端蜂窝状网络结构示意如图 1-7 所示。

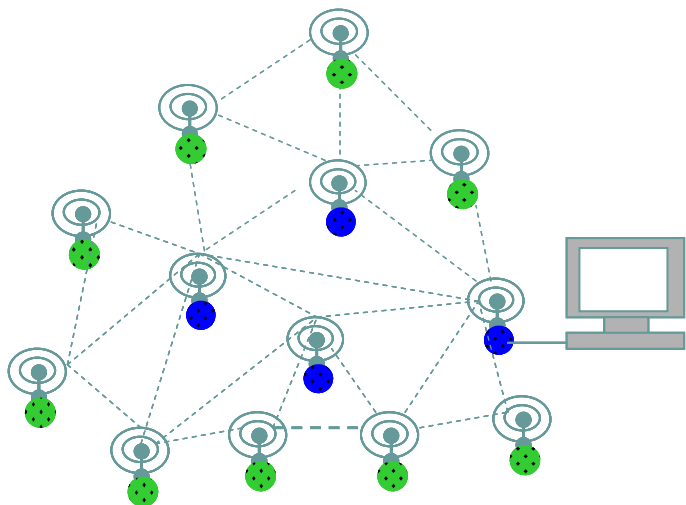


图 1-6 无线数据终端网状网络连接结构示意图

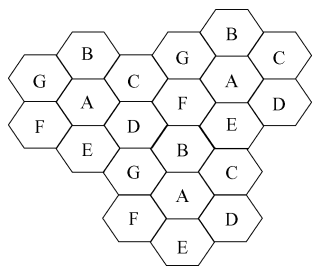


图 1-7 移动数据终端蜂窝状网络结构示意图

现在随着计算机和通信技术的不断发展，模拟通信网已经萎缩，数字通信网停滞不前，数据通信网络不断壮大，如短消息业务、无线 Internet 业务、移动定位业务、移动电子商务、移动多媒体业务等。

## 1.2 无线数据通信方式与类型

### 1.2.1 无线数据通信与计算机通信、数字通信

数据通信是指两个数据终端 (DTE) 之间的通信。无线数据通信是指两个数据终端 (DTE) 之间的无线通信。计算机属于智能化程度较高的数据终端，因此计算机通信应归入数据通信的范畴。

从概念上讲，数据通信应包含计算机通信。由于计算机是目前应用最普遍的数据终端，有许多人又将数据通信与计算机通信等同起来，因此，在许多地方数据通信与计算机通信几乎成了同义语。狭义地讲，数据通信仅指计算机通信中的通信子网的具体实现，它完成通信协议中的下三层功能，主要解决两个数据终端之间的通信传输问题，而计算机通信着重于数据信息的交互，即更侧重于计算机内部进程之间的通信。

从电信的角度来讲，数据通信是一种新的业务。数据通信完全不同于现在的电话通信，

它具有许多新的概念和思想，是一种新的通信技术。简单地说，数据通信是计算机应用开发的产物。由于计算机的大量存在，而单个计算机在使用中不能充分地发挥其潜力。人们自然地想到用通信线路把计算机连接起来进行远程通信，实现资源共享，于是出现了数据通信。从这个角度讲，由于计算机的普遍使用，许多信息以数据形式存在，如何传递数据信息也就是数据通信需要考虑的内容。

数字通信是电信号的一种传输方式，传输出“0”和“1”组成的数字码流，这些码流既可以表示成数据信息，也可以代表语音和图像信息等，它与模拟通信相对应，主要解决模拟信号的数字化传输问题。一般来说，数字通信并不针对某种用户业务，因而不涉及用户终端。但数据通信却是要针对数据业务的。数据既可以通过调制技术在模拟通信系统上传输，也可以在数字通信系统上传输。也就是说，无线数据通信所使用的信道既可以是模拟信道，也可以是数字信道。

### 1.2.2 无线数据通信的工作方式

无线数据通信的工作方式可分为单工通信、半双工通信及全双工通信。

#### 1. 单工通信工作方式

所谓单工通信，是指数据消息只能单方向进行传输的一种通信工作方式，如图 1-8（a）所示，发送端只管发送，接收端只管接收。

#### 2. 半双工通信工作方式

所谓半双工通信方式，是通信双方使用同一个信道或同一根传输线，既作输入又作输出，虽然数据可以在两个方向上传送，但通信双方不能同时收发数据的形式，如图 1-8（b）所示。

采用半双工时，通信系统每一端的发送器和接收器，通过收/发开关接到通信线上，由软件控制的电子开关进行方向的切换，因此会产生时间延迟。

目前，大多数终端和串行接口都为半双工模式提供了换向功能，也为全双工模式提供了两条独立的引脚。在实际使用时，一般并不需要通信双方同时既发送又接收，像打印机这类的单向专送设备，半双工就能胜任，也无须双向。

#### 3. 全双工通信工作方式

所谓全双工通信，是指通信双方可同时进行双向传输消息的工作方式，如图 1-8（c）所示。

当数据的发送和接收分流，分别用不同的信道或两根不同的传输线传输时，例如采用频分复用或时分复用技术，通信双方都能同时进行发送和接收操作，此传送方式就是全双工模式。

在全双工方式下，通信系统的每一端都设置了发送器和接收器，因此，能控制数据同时在两个方向上传送，即向对方发送数据的同时，也可以接收对方送来的数据。全双工方式无须进行方向的切换，因此这种传送方式对那些不能有时间延迟的交互式应用（如远程监测和控制系统）十分有利。

### 1.2.3 无线数据通信的信号传输方式

无线数据通信的信号传输方式，是指数据信号在信道上的传输时所采取的方式。根据一

次传输数据的多少可将数据传输方式分为并行传输和串行传输，按照收发两端信号同步状态分为同步传输与异步传输。

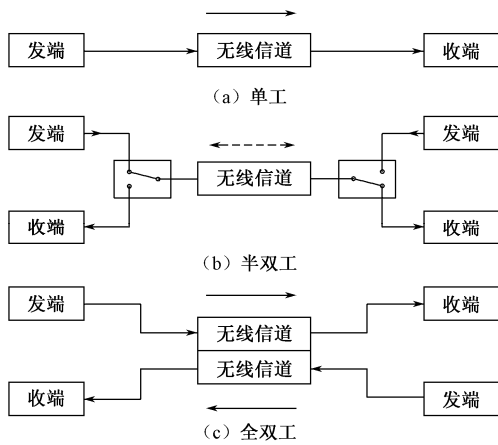


图 1-8 无线数据通信的工作方式示意图

1. 数据信号并行传输方式

采用并行传输时，多个数据位同时在通信设备间的多条通道（信道）上传送，并且每个数据位都有自己专用的传输通道。

图 1-9 描述了通信设备之间具有多条传输通道时的并行传输情况。例如，采用 7 单位代码字符时，可以用 8 条信道并行传输，另加 1 条“选通”线用来通知接收器，以指示各条信道上已出现某一字符的数据信息。

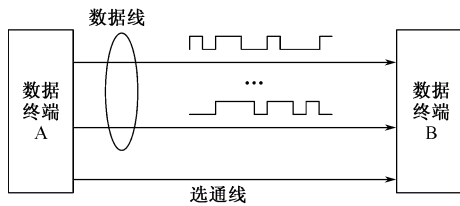


图 1-9 数据信号并行传输方式

这种传输方式的特点是数据传输速率相对较快，不需要额外措施就实现了收发双方的字符同步。其缺点是需要传输线路（信道）多，设备复杂、成本高，故无线数据通信较少采用，一般适用于计算机内部和其他高速数据的近距离传输。在计算机中，CPU 和 RAM 之间、计算机和打印机之间都是并行传输的应用。

2. 数据信号串行传输方式

串行传输是数据码流以串行方式在一条信道上，按照顺序一位一位地在通信设备之间的上传输，如图 1-10 所示。

因为，在数据终端内部各个部件之间，一般以并行方式传输数据，所以数据传至传输线路之前，需要先被送入发送端的并/串变换器中，通过变换器，数据逐位传至线路上到达接收端时，数据先进入接收端的串/并变换器中，使数据传输从串行传输变成并行传输。

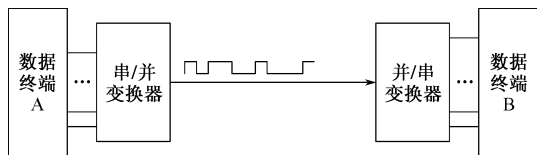


图 1-10 数据信号串行传输方式

该方法优点是传输线路（信道）少，易于实现，缺点是为了解决收、发端双方码组或字符同步，需外加同步措施。通常，无线数据通信和远距离有线数据传输时采用较多。例如 GPRS、计算机 RS-232 串行接口输入输出、USB 通用串行总线输入/输出等，均采用串行传输方式。

### 3. 数据信号同步传输与异步传输

通过上面讨论可以看出，并行数据传输的同步问题比较简单，通过在收、发两端之间多加一根控制线就可完成数据的同步（步调一致）。

在串行传输时，收发两端只有一条线路（信道），在此信道上既要完成一位一位的数据流传输，同时，还要确定数据字符步调问题，所以要采取特殊措施，才能准确地传送信息。

根据串行传输实现字符同步方式的不同，数据传输分为异步传输和同步传输两种方式。

#### 1) 异步传输

异步传输方式是指收、发两端各自有相互独立的位（码元）定时时钟，数据率是收发双方约定的，收端利用数据本身来进行同步的传输方式。

异步传输中，每次传送一个数据信息字符代码，即在发送的每一个字符代码的前面均加上 1 个“起”信号，其长度规定为 1 个码元，极性为“0”，后面均加 1 个“止”信号。对于国际电报 2 号码，“止”信号长度为 1.5 个码元，对于国际 5 号码或其他代码，“止”信号长度为 1 个或 2 个码元，极性为“1”。数据信息字符可以连续发送，也可以单独发送；不发送信息字符时，连续发送“止”信号。

因此，每一数据信息字符的发送的起始时刻可以是任意的（这正是称为异步传输的含义），但在同一个字符内各码元长度相等。这样，接收端可根据字符之间的从“止”信号到“起”信号的跳变（“1” — “0”）来检测识别一个新数据信息字符的“起”信号，从而正确地区分一个个数据信息字符。因此，这样的数据信息字符同步方法又称为起止式同步。

异步传输的优点是实现简单，不需要收、发两端之间的同步专线，即收发双方的时钟信号不用要精确同步。缺点是每个数据信息字符增加了起、止的比特位，降低了传输效率，所以常用于 1 200 b/s 及其以下的低速数据传输。图 1-11 表示了异步传输的情形。

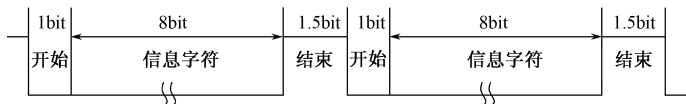


图 1-11 数据信号异步传输方式示意图

#### 2) 同步传输

同步传输是相对于异步传输而言的，指收发双方采用统一的时钟节拍来完成码元同步的



基础之上实现数据信息字符的传送。即在数据传输中，必须建立位定时同步和帧同步。

位定时同步又称比特同步，其作用是使数据电路终端设备接收端的位定时时钟信号和 DCE 收到的输入信号同步，以便 DCE 从接收的信息流中正确识别一个个信号码元，产生接收数据序列。

帧同步又称群同步，一群或一串数据信息为一帧，其中每帧的开头和结束加上预先规定的起始序列和终止序列作为标志，这些特殊序列的形式决定于所采用的传输控制规程。

在 ASCII 代码中，用 SYN（码型为“0010110”）作为“同步字符”，通知接收设备表示一帧的开始；用 EOT（码型为“0000100”）作为“传输结束”字符，以表示一帧的结束。同步传输的数据格式如图 1-12 所示。

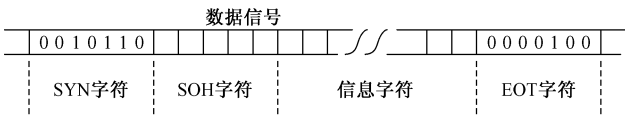


图 1-12 数据信号同步传输格式示意图

与异步传输相比，同步传输在技术上要复杂一些，但它不需要对每个数据字符单独加起、止码，只是在一群数据信息字符的前后加上标志序列，因此传输效率较高。

同步传输通常用于速率为 2 400 b/s 及其以上的数据传输。

## 1.2.4 无线数据通信的过程与信息流的方式

### 1. 数据通信的过程

数据通信网络一般应包括：信源（产生数据的数据源）、信宿（接收数据的数据宿）、信道、通信控制器以及并/串变换器、串/并变换器等，其数据通信的一般过程如下：

- （1）当信源要发送信息时，根据通信规程，应先通过通信控制器取得信道的使用权。
- （2）信源将要发送的信息并行传送给通信控制器。
- （3）通信控制器根据发送端的要求，把信息分为报文组或帧，将其串行送入变换器。
- （4）变换器对初始信号进行必要的转换，再串行送入信道进行传输。
- （5）接收端的变换器将传输信号恢复为原始信号，送往通信控制器，并判断信息是否是发给自己的信息。如果是自己的信息，通信控制器把收到的串行数据格式转换成并行数据格式，并存入自己的缓冲器中，当所有的报文收齐后，再并行地发送给信宿；如果不是，则不做任何处理。

### 2. 数据通信的信息流的方式

数据通信按信息流方式可以分为如下的基本类型。

#### 1) 数据处理/查询系统

这种类型的数据通信系统的信息流如图 1-13 所示。

在中央处理单元（CPU）的文件中存在有可查阅的大量数据，当数据终端查询时，终端首先与 CPU 建立数据链路，然后发送查询命令；CPU 收到查询命令（输入数据）进行检查，

根据检查结果调出相应的程序和数据进行处理，并将处理结果进行必要的编辑以适应线路传送和终端接收的形式；最后发送回终端，作为对查询的响应。例如飞机订票系统、银行系统和信息检索系统就属于此种类型。

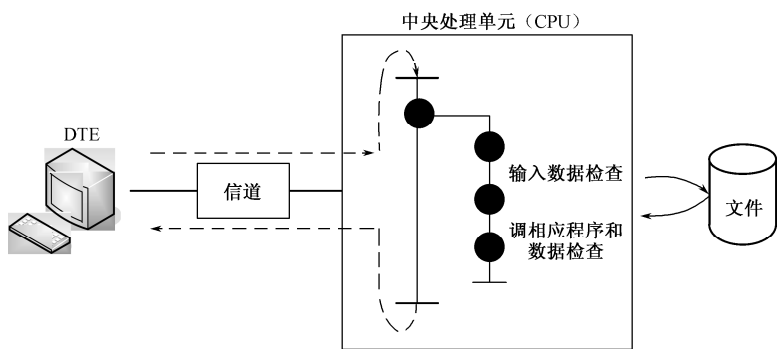


图 1-13 数据处理/查询系统的信息流示意图

2) 信息交换系统

信息交换系统的信息流如图 1-14 所示。

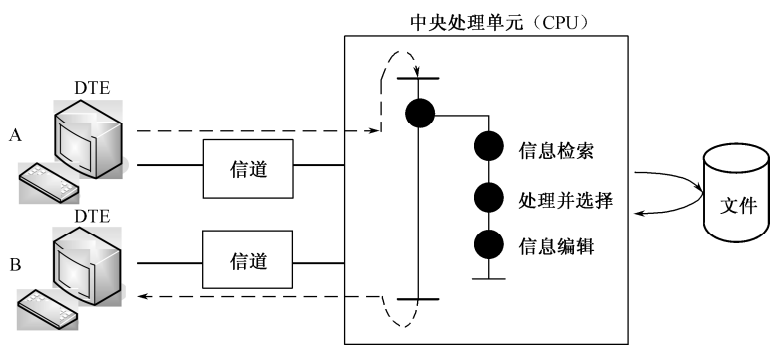


图 1-14 数据信息交换系统的信息流示意图

若终端 A 需要将信息送到终端 B，终端 A 首先建立与 CPU 的数据链路，并将要交换的信息送到 CPU；CPU 收到该信息后对其进行检查和处理，并选择所需要的目的地终端 B；然后按照接收终端对信息格式的要求，对交换信息进行必要的编辑，并建立与目的地终端 B 的数据链路；将信息发送给终端 B，完成信息交换。例如，票证交换系统就是一种信息交换系统。

3) 数据收集和分配系统

数据收集和分配系统的信息流如图 1-15 所示。

作为数据收集系统，从很多数据终端发送来的数据被 CPU 收集，收集的数据被存入文件中，以备进一步处理，如气象观察系统。这种系统可以作为分配系统。

在实际的数据通信系统中，这些形式是组合在一起使用的，可以提供更广泛的业务。

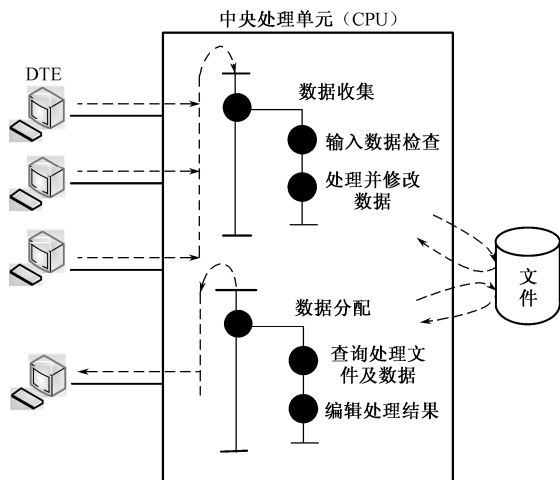


图 1-15 数据收集和分配系统的信息流示意图

## 1.2.5 数据通信的业务类型

数据通信技术的发展是离不开它所支持和提供的业务的。从信息载体的角度说，数据业务可以有多种分类。

按照传送速率，可分为低速、中速和高速数据业务。

按照提供的带宽，可分为窄带、宽带和超宽带数据业务。

按照数据通信业务是否增值，可分为基础业务和增值业务，在工信部发布的《电信条例》中，数据业务分为基础数据业务和增值数据业务两大类。

### 1. 基础数据业务

基础数据业务主要指公共数据传送业务和移动数据业务。

公共数据传送业务是利用电路交换、分组交换或租用电路组成的固定公共数据通信网上开发的以传送数据为目的的业务。

按照所用技术的不同，公共数据传送业务包括分组交换、数字数据网（DDN）、综合业务数字网（ISDN）、帧中继、异步传送方式（ATM）和 IP 业务等，其中分组交换、帧中继和 ATM 业务都采用面向连接的分组交换技术，具有统计复用、用户共享网络带宽等功能，仅它们所用的通信协议、能提供的接入速率、控制能力和综合能力有所不同。它们部分有基本业务和用户选用的业务。基本业务是指向所有网上的用户提供的基本服务功能，包括永久虚电路（PVC）和交换虚电路（SVC）业务。用户选用业务是为了满足用户特殊要求而向用户提供的特殊业务功能。

利用公用陆地移动蜂窝通信网作为承载网提供的数据业务称为移动数据业务。具体可以包括短消息业务、速率可达 64 kb/s 的中速移动数据业务、速率在 128 kb/s 以上的移动多媒体业务。移动数据业务也可以采用电路交换和分组交换方式来实现。

中国电信为提供上述数据传送业务，先后建设了覆盖全国的各种数据通信网络，包括中国公用计算机互联网（CHINANET）、中国公众多媒体通信网（CNINFO）、中国公用分组交换数据网（CHINAPAC）、中国公用数字数据网（CHINADDN）、中国公用帧中继宽带业务网

(CHINAFRN) 和移动数据网。

## 2. 增值数据业务

增值数据业务的概念最开始就是从数据业务引入的。它是指在原基础网络设施的基础上增加必要的设备构成增值网后，向用户提供新的业务，大大提高原基础网络设施的使用价值。在公共数据网（不包括互联网）上开发的增值业务很多，主要有电子邮件（E-mail）、可视图文（Videotext）、电子数据交换（EDI）、传真存储转发（S/F Fax）、在线信息库存储和检索以及在线数据处理和交易处理等。由于 Internet 的广泛应用，上面提到的电子邮件、可视图文和电子数据交换这些增值业务逐渐被 Internet 上的类似业务所取代了。

## 1.3 无线数据通信的主要性能指标

数据通信系统同模拟通信系统、数字通信系统一样具有一些技术性能指标，其中有些概念与后两种通信系统类似。为了知识内容的完整性和系统性，我们在这里对数据通信系统的主要技术性能指标完整地给予介绍。

### 1. 无线信道带宽

带宽有信道带宽和信号带宽之分，一个信道能够传送电磁波的有效频率范围就叫该信道的带宽，这主要由信道的特性决定。对信号而言，信号所占据的频率范围就是信号的带宽。

无线信道带宽除了与信道的特性有关外，还要由国际频管组织的频段划分与通信体制有关。

通常要求信号带宽小于信道带宽，但对于无线数据通信来说，奈奎斯特信道容量定理决定了数据信号的最大带宽。

### 2. 信号传播速度

信号传播速度是指信号在信道上每秒钟传送的距离，单位是 m/s。由于我们所用的通信信号都是以电磁波的形式出现，因此信号传播速度略低于光在真空中的速度，基本上是  $3 \times 10^8$  m/s。

信号传播速度  $v$  与信号波长  $\lambda$  和频率  $f$  的关系是： $v = f\lambda$ 。随着传输介质的不同，可能会有少许变化。一般来说，信号传播速度是个常量，介绍它的目的主要是和“数据传输速率”的概念加以对比。

### 3. 数据传输速率

#### 1) 数据码元传输速率（波特率）

数据码元传输速率简称传码率，表示单位时间（每秒）内传输的数据码元（符号）个数，单位为波特（Baud，简写为 B），故也称波特率。

例如周期为  $T_0$  的四电平脉冲数据信号（图 1-16），码元出现的频率为  $f_0 = f_b = 1/T_0$ ，则传码率  $R_B = f_b = 1/T_0$ ， $R_B$  只与码元宽度有关，与进制无关。

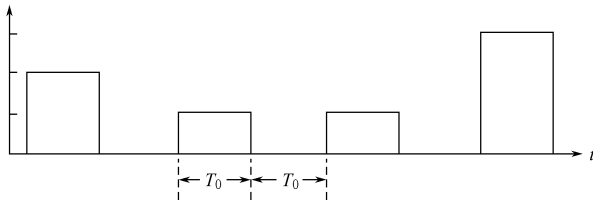


图 1-16 四电平脉冲数据信号波形示意图

## 2) 数据信息传输速率 (比特率)

数据信息传输速率简称传信率, 也称比特传输速率, 表示单位时间 (每秒) 内传送数据信息的比特数, 单位为 b/s。

二进制数据传输时, 码元传输速率和信息速率一致。

对  $N$  进制传输时, 其传码速率  $R_B$  和传信率  $R_b$  可以通过下述公式来换算:

$$R_b = R_B \log_2 N \quad (1-3)$$

注意, 在描述数据传输能力时, 通常使用传信率。

## 3) 最大传输速率

每个信道传输数据的速率有一个上限, 我们把这个速率上限叫做信道的最大传输速率, 也就是信道容量。

## 4. 吞吐量

吞吐量是信道在单位时间内成功传输的信息量, 单位一般为 b/s。

例如某信道 10 min 内成功传输了 8.4 MB 的数据, 那么它的吞吐量就是  $8.4 \text{ MB} / 600 \text{ s} = 14 \text{ kb/s}$ 。注意, 由于传输过程中出错或丢失数据造成重传的信息量, 不计在成功传输的信息量之内。

## 5. 数据延迟

数据延迟指从发送者发送第一位数据开始, 到接收者成功地收到最后一位数据为止, 所经历的时间。它又主要分为传输延迟、传播延迟两种, 传输延迟与数据传输速率和发送机/接收机以及中继和交换设备的处理速度有关, 传播延迟与传播距离有关。

## 6. 频带利用率

数据信号的传输需要一定的频带。数据传输系统占用的频带越宽, 传输数据信息的能力越大。因此, 在比较不同数据传输系统的效率时, 只考虑它们的数据传信速率是不充分的。因为, 即使两个数据传输系统的传信速率相同, 它们的通信效率也可能不同, 这还要看传输相同信息所占的频带宽度。故真正衡量数据传输系统有效性的指标是单位频带内的传输速率, 即频带利用率。

数据通信系统频带利用率  $\eta$  有两种表示方法, 一是在传码率意义下的频带利用率, 一是在传信率意义下的频带利用率, 常用的是式 (1-4) 传码率意义下的频带利用率。

$$\eta = \frac{\text{传码率}}{\text{频带宽度}} \quad (1-4)$$

$$\eta = \frac{\text{信息速度}}{\text{频带宽度}} \quad (1-5)$$

## 7. 差错率

差错率包括比特差错率、码元差错率与分组差错率。差错率是衡量通信信道可靠性的重要指标，在数据通信中最常用的是比特差错率和分组差错率。

### 1) 误信率（比特差错率）

在传输过程中发生错误信息比特数与传输的总信息比特数之比叫误信率，或称比特差错率。

$$P_b = \frac{\text{错误信息数}}{\text{传输的总信息数}} \quad (1-6)$$

### 2) 误码率（码元差错率）

在传输过程中发生误码的码元个数与传输的总码元数之比叫误码率，或称码元差错率。

$$P_e = \frac{\text{发生误码个数}}{\text{传输的总码数}} \quad (1-7)$$

### 3) 分组差错率

在传输过程中发生误码的分组个数与传输的总分组数之比称分组差错率。

## 1.4 无线数据通信研究的主要内容与发展目标

### 1. 无线数据通信研究的主要内容

无线数据通信研究得主要包括数据的无线信道上传输、数据的计算机处理、数据集集中、数据交换、差错控制和传输规程等。目的是为了使数据更有效、更可靠地传输。

由无线数据通信的特点可知，无线数据通信与其他通信方式所包涵的基本内容有很大的区别。特别是由于信道特点、信息载体的不同和具有复杂的协议，使无线数据通信所研究的内容也有很大的不同。与此同时还有一些全新的概念，这些概念在学习无线数据通信中要引起足够的重视。

#### 1) 无线数据传输

无线数据传输主要负责为数据提供一个可靠而有效的无线传输通路，研究适合无线信道传输的数据电信号形式，即数据压缩编码与译码、加密与解密、调制与解调、干扰与抗干扰、多路复用技术等技术，以及构成传输媒体和用来控制电信号的各种传输设备。

#### 2) 通信接口

通信接口负责把发送端产生的信号转换为适合在信道中传输的信号，并且把传送到接收端的信号变换为终端可接收的形式。

数据通信系统根据不同的应用要求，规定了不同类型的接口标准，有国际标准、国家标准，也有公司自己制定的标准。但对开放性的用户接口通常是采用国际标准或国家标准，以利于互联互通。关于通信接口的详细内容将在后续章节中介绍。

### 3) 数据交换

数据交换是指网络中各节点之间的信息交互方式。

交换是数据通信的重要内容，主要包括数据交换的原理与概念。数据交换是为了解决传输资源共享。

数据交换的主要方式有多种：电路交换、报文交换、分组交换、帧中继、ATM 等，其中分组交换在实际的数据通信网中较多采用。

在一个采用分组交换的数据通信网中，除了在相邻交换节点之间实现数据传输与数据链路控制规程所要求的各项功能外，在每个交换节点上还需完成分组的存储与转发、路由选择、流量控制、拥塞控制、用户入网连接以及有关网络维护、管理等多方面的工作。

### 4) 通信处理

通信处理是数据通信中最复杂的部分，涉及数据的差错控制、格式化处理、速度转换、码型转换、流量控制、寻址、路由选择等内容。

### 5) 通信协议

数据通信协议也即传输控制规程，通信协议是数据通信规则的集合，是双方为准备有效地进行通信所必须遵循的规程和约定，是通信网络的“大脑”，它与网络操作系统、网络管理软件共同控制和管理着数据网络的运行。

数据通信协议可以分为两类：一类是与数据通信网有关的协议，包括网内节点与节点间，以及网络与端系统间的协议；另一类是端系统与端系统之间的协议，它们是在前一类协议所实现的基础上，为了实现端系统间的互通与达到一定的应用目的所必需的协议。这部分内容将在第 11 章中介绍。

### 6) 多路复用

解决如何把多路信号组合起来在一条物理信道上同时传输，以提高通信效率。常见的复用方式有频分复用、时分复用、波分复用、码分复用等。

### 7) 同步

同步是数字化信息正确传输的前提，可分为载波同步、位同步、群同步、网同步等。

## 2. 无线数据通信技术的发展目标

任何人在任何时间和任何地点都可以和世界上的任何人进行通信的理想境界是无线数据通信追求的目标。

无线数据通信技术的发展目标主要有以下几个方面：

(1) 现代无线数据通信技术应能为用户提供更安全和保密、环境适应能力与抗干扰能力更强的数据通信服务。

(2) 现代无线数据通信技术应能为用户提供更大的吞吐量、更高的传输速率、更低的延迟, 以实现更快的通信网络的运行。

(3) 现代无线数据通信技术向着数字处理技术的开发应用发展, 使无线和有线通信的数字化, 通信设备小型化、智能化。

(4) 现代无线数据通信技术要能够提供网络硬件与软件之间的“无缝”连接, 以及网络之间的“无缝”连接。这样, 通信的最终用户意识不到用户业务是通过不同网络上的不同供应商提供的设备传送的。

(5) 现代无线数据通信技术要支持任意类型的应用, 它不仅能支持现有的各种数据通信业务, 而且也能支持未来可能出现的通信新业务。

(6) 开发一系列能在不同通信设备与网络上运行的技术规程和标准。现代无线数据通信技术都必须支持大容量的网络, 而现有的一些网络都包含有不同商家的不同产品, 且网络运行的软件大都工作在不同的载体上。因此, 制定一个统一的规程和标准已是无线数据通信界共同的目标。

(7) 从网络供应商的角度看, 现代无线数据通信技术必须提供更多更好的网络管理工具。这些工具可以使网络营运商能实时、详细地监控网络设施, 以便为用户提供可靠的服务。

当然, 以上无疑是无线数据通信的极高目标, 一些具体通信技术不可能同时满足。

## 思考与练习

1-1 试说明基带信号与频带信号, 模拟信号、数字信号与数据信号的区别。

1-2 试分别画出模拟、数字和数据信号通信系统的模型, 并简要说明模型中各部分的作用。

1-3 衡量数据通信系统的主要性能指标有哪些?

1-4 某数据源的符号集由 A、B、C、D、E、F 组成, 设每个信号独立出现, 概率分别为  $1/4, 1/4, 1/16, 1/8, 1/16, 1/4$ , 则信源输出符号平均信息量大小?

1-5 设数据传输系统传送二进制信号, 码元传输率  $R_B=2\ 400\text{ B}$ , 该系统的信息速率为多少? 若该系统改为十六进制信号, 码元速率不变, 此时的信息速率为多少?

1-6 在强干扰环境下, 某电台在  $5\text{ min}$  内共收到正确信息量为  $355\text{ MB}$ , 假定系统信息速率为  $1\ 200\text{ kb/s}$ , 系统的误信率为多少?

1-7 某系统经长期测定, 其误码率为  $P_e=10^{-5}$ , 系统码元速率为  $1\ 200\text{ B}$ , 在多长时间可收到 360 个错误码元?

1-8 数字通信系统的优点有哪些?

1-9 试述码元速率、信息量、信息速率的定义、单位和符号, 并说明二进制和多进制时码元速率和信息速率的相互关系。

1-10 如何衡量数据通信的频带利用率?

1-11 什么是单工、半双工、全双工数据传输?

1-12 设带宽为  $3\ 000\text{ Hz}$  的模拟信道, 只存在加性高斯白噪声, 如信号噪声功率比为  $30\text{ dB}$ , 试求这一信道的信道容量。

1-13 异步传输中, 假设停止位为 1 位, 奇偶校验位为 1 位, 数据位为 7 位, 求传输效率。



# 第 2 章 无线通信信道

任何一个通信系统与网络，从大的方面均可视为由发送端、信道和接收端三大部分组成。因此，无线信道是无线数据通信系统与网络必不可少的组成部分，无线信道特性的好坏直接影响系统与网络的总特性。

## 2.1 无线信道的概念与分类

### 1. 无线信道的概念

信道的应用参考点不同，信道的概念及定义就不同，狭义的无线信道概念，是指以电磁波、红外线、可见光波等载体，在空间或宇宙中构成的信号物理通路，或者说，是指空间或宇宙中电波通过的一段频带。

无线信道是非导向传输信道，在空间或宇宙中自由直线方式传播，传播速度为  $3\times10^8$  m/s，同时具有折射、反射、绕射和干涉等波的特性。（导向传输信道，如光缆、电缆、双绞线等，光波或电磁波被导向沿此固体信道传播。）

无线信道为开放性信道，时时刻刻受到人为或自然现象的影响与干扰，传输特性不如有线信道稳定和可靠，但无线信道具有方便、灵活，通信者可移动等优点，在现代生活中，越来越受到人们的重视。

### 2. 无线信道的分类

无线信道的分类方法很多，依照无线电传输特性不同，无线信道可分为近地空间信道、对流层散射信道、电离层折射信道、流星余迹反射信道、卫星通信信道等。对应的通信方式为地面视距通信、近地空超视距通信、对流层散射通信、电离层反射通信、卫星通信等。

依照无线信道中传输的电波频率或波长不同，可分为极低频、超低频、特低频、…、至高频、光波频段等，对应的波长为极长波、超长波、特长波…、丝米波、光波等，具体见表 2-1 所示。不同频率范围对应的传输媒质与用途见表 2-2 所示。

表 2-1 频段名称、频率范围

| 序号 | 频段名称      | 频率范围<br>(含上限，不含下限) | 波段名称 | 波长范围<br>(含下限，不含上限) |
|----|-----------|--------------------|------|--------------------|
| 1  | 极低频       | 3~30 Hz            | 极长波  | 100~10 Mm          |
| 2  | 超低频       | 30~300 Hz          | 超长波  | 10~1 Mm            |
| 3  | 特低频       | 300~3000 Hz        | 特长波  | 1 000~100 km       |
| 4  | 甚低频 (VLF) | 3~30 kHz           | 甚长波  | 100~10 km          |
| 5  | 低频 (LF)   | 30~300 kHz         | 长 波  | 10~1 km            |
| 6  | 中频 (MF)   | 300~3000 kHz       | 中 波  | 1 000~1 m          |
| 7  | 高频 (HF)   | 3~30 MHz           | 短 波  | 100~10 m           |

续表

| 序号 | 频段名称      | 频率范围<br>(含上限, 不含下限)                  | 波段名称 |    | 波长范围<br>(含下限, 不含上限) |
|----|-----------|--------------------------------------|------|----|---------------------|
| 8  | 甚高频 (VHF) | 30~300 MHz                           | 米 波  |    | 10~1 m              |
| 9  | 特高频 (SHF) | 300~3000 MHz                         | 分米波  | 微波 | 10~1 dm             |
| 10 | 超高频 (SHF) | 3~30 GHz                             | 厘米波  |    | 10~1 cm             |
| 11 | 极高频 (EHF) | 30~300 GHz                           | 毫米波  |    | 10~1 mm             |
| 12 | 至高频 (THF) | 300~3000 GHz                         | 丝米波  |    | 1~0.1 mm            |
| 13 | 光波频段      | 10 <sup>5</sup> ~10 <sup>7</sup> GHz | 光波   |    |                     |

表 2-2 频率范围、传输媒质与用途

| 频 率 范 围            | 传 输 媒 质                     | 用 途                                |
|--------------------|-----------------------------|------------------------------------|
| 3 Hz~30 kHz        | 对称型电话线（缆）、地面、水下等            | 音频、电话、数据终端、长距离导航等                  |
| 30~300 kHz         | 对称型电话线（缆）、地面等               | 导航、信标频、率标准、电力线通信等                  |
| 300 kHz~3 MHz      | 对称型电话线（缆）、同轴电缆、地面视距等        | 导航、中波调幅广播、移动陆地通信、业余无线电等            |
| 3~30 MHz           | 双绞型线（缆）、同轴电缆、电离层反射等         | 导航、移动无线电话、短波广播、短波军事通信、业余无线电等       |
| 30~300 MHz         | 双绞型线（缆）、同轴电缆、波导、地面视距、空中超视距等 | 导航、电视、调频广播、空中管制、车辆通信、导航、集群通信、无线寻呼等 |
| 300 MHz~3 GHz      | 波导、地面视距、空中超视距等              | 电视、空间遥测、导航、点对点通信、移动通信等             |
| 3~30 GHz           | 波导、地面视距、空中超视距、宇宙空间等         | 微波接力、卫星和空间通信、导航等                   |
| 30~300 GHz         | 波导、空中超视距、宇宙空间等              | 导航、微波接力、射电天文学等                     |
| 300~3 000 GHz      | 波导、宇宙空间等                    | 导航、射电天文学等                          |
| $10^5\sim10^7$ GHz | 光纤、宇宙空间等                    | 光纤（缆）通信、宇宙空间激光通信、空中激光通信等           |

2.2 电波传播方式与无线信道特点

无线电波的传播分直射、反射、折射、绕射和散射方式（现象），传输信道分近地空间信道、对流层散射信道、电离层折射信道、卫星通信信道等，不同信道的电波直射、反射、折射、绕射和散射特性不同，下面首先介绍无线电波的传播方式，然后介绍不同信道的传输特点。

2.2.1 无线电波的传播方式

无线电波的传播分直射、反射、透射、折射、绕射、散射等方式或现象。

1. 无线电波直射

根据无线电波的特性，在传输距离内无遮挡的均匀大气媒质中，应以恒定的速度沿直线传播，由于能量的扩散与媒质的吸收，传输距离越远信号强度越小，并且大气吸收衰减与电波频率有关，频率越高，衰减越大，如图 2-1 所示。

在非均匀的大气媒质中传播时，速度会发生变化，同时还会产生反射、折射、绕射、散射的现象，其传播能量一般会小于直射波。

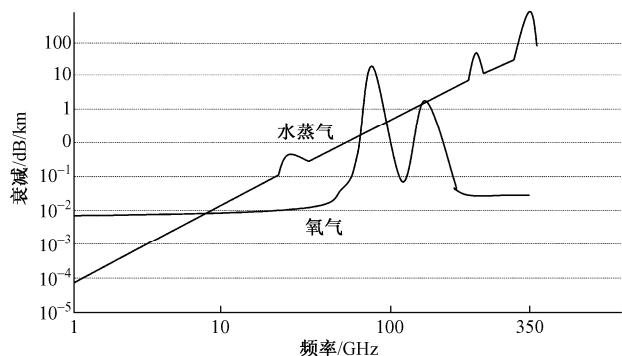


图 2-1 大气吸收损耗（衰减）

## 2. 无线电波反射

当电波碰到的建筑物或其他障碍物大于其波长时，会发生反射，如图 2-2 所示，反射波的强度比直射波差，但多重反射会形成多条传播路径到达接收端，故又称多径反射波，会造成多径衰落。

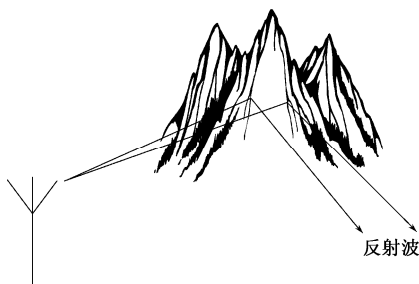


图 2-2 无线电波反射

反射波是当无线电波传播途中遇到比波长尺寸大得多的物体时发生，理论分析时由简到繁，分为无限大平面边界、有限大平面边界、曲面边界、非理想曲面边界几个阶段进行分析。

## 3. 无线电波折射与透射

当电波由一个媒质进入到另一个媒质，例如电波进入水下、地下、大气层或墙壁等，由于媒质的密度、介电常数等不同，传播速度与路径偏转不同，这种现象称为无线电波折射，如图 2-3 所示，并将电波由一个媒质穿过另一个媒质，例电波穿过墙壁、地板、家具，或电波穿过大气层进入外侧空间等时，称为无线电波透射。

电波的折射强度与入射波的波长、强度、角度和障碍物的透射率等因素有关，并且一般折射强度比直射波差。

## 4. 无线电波绕射

当电波遇到较大的障碍物（山丘、建筑物等）时，会通过边缘绕到其背后继续传播，称为无线电波绕射，如图 2-4 所示。绕射后到达接收点的传播信号，其强度与反射波相当，并且波长越长，绕射能力越强，但是当障碍物尺寸远大于电波波长时，绕射就变得微弱。

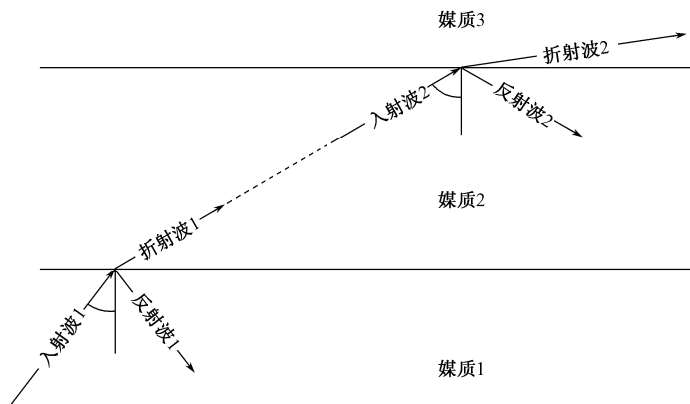


图 2-3 无线电波折射与透射

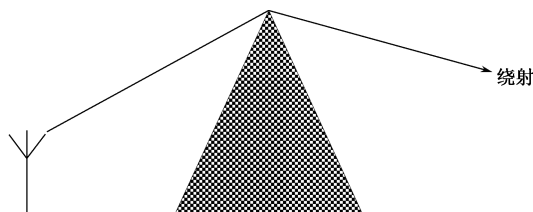


图 2-4 无线电波绕射

## 5. 无线电波散射

当电波遇到粗糙障碍物或小物体，如雨点、树叶、微尘等，会产生大量的杂乱无章的反射，即漫反射，称为无线电波散射，对流层散射传播如图 2-5 所示。散射造成能量的分散，形成电波的损耗更大，到达接收点的传播信号强度最弱。

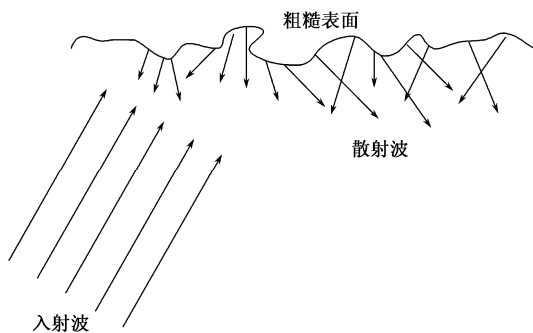


图 2-5 无线电波绕射

根据无线信道特点不同分为，近地空间传播信道、对流层散射信道、电离层折射信道、宇宙及外层空间卫星通信信道的特点。

### 2.2.2 近地空间信道的特点

无线电波近地空间信道，包括电波地表绕射、近地空间直射和地面反射，地面反射又可

分地表反射与地面物体反射，如图 2-6 所示。

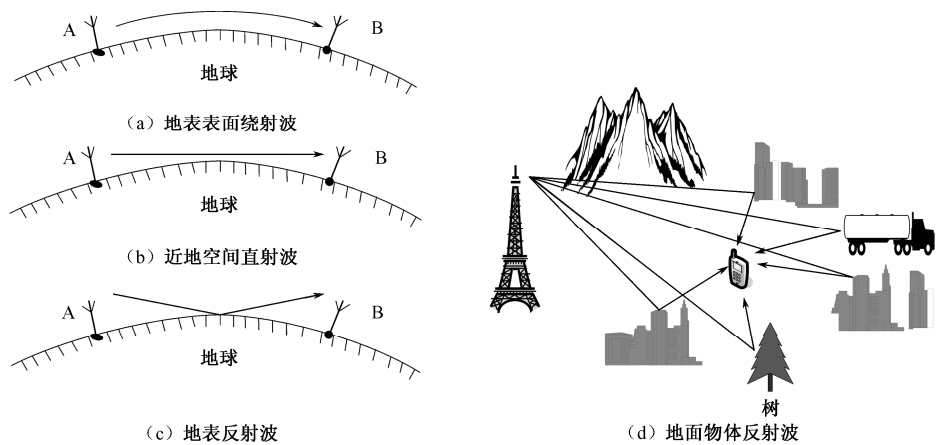


图 2-6 电波近地空间传播示意图

近地空间传播主要特性：

(1) 地表绕射波是利用天线与大地之间的静电感应进行电波传播，只能沿着地球表面传播的一种电波传播方式，由于大地的传导率将水平极化波短路，因此地波为垂直极化波，如图 2-6 (a) 所示，空箭头表示波的传播方向。

大地对绕射波吸收损耗很大，并于地表的导电性和电波频率等有关。导电性越好地阻越小，吸收越小，电波传播损耗越小。地表传播特性好的地面依次是海面，水面（淡水），湿土地面，干燥地面。电波频率越低，损耗越小，波长越长绕射能力越强。

当频率合适且功率足够时，地波传播可围绕地球进行通信。2 MHz 以下地表绕射波传播基本不受天气及太阳辐射等外界因素影响。

地表绕射多用于海上通信、海岸电台与船舶电台之间的通信及近距离的陆地无线电通信，并需要较高的发射功率。

(2) 近地空间直射波，受地球曲率和天线高度的影响，地面上两点间的直线通信距离，一般小于 60 km 的视距范围。若近地空间超视距直线通信，只能在地面与空中、空中与空中之间完成，工作频率一般在 30 MHz 以上。

虽然近地空间直射波的可以使用不同的工作频率，但近地空间的多径衰落、邻台干扰、大气干扰、同频干扰（由同频天波引起的多径）等影响非常严重，目前近地空间直射波多用于微波接力通信、微波移动通信和地面与空中超视距通信等。

(3) 地表反射波与地面物体反射波，是近地空间传播存在的主要问题之一，其信号强度比直射波弱，并带来的多径衰落会直接影响通信的正常进行，所以要采取抗多径衰落的分集接收技术或尽量避免其存在。

(4) 无线电波在近地空间传播时，如果在传播路径中遇到起伏的地形、建筑物和高大的树叶等障碍物，就会在障碍物的后面形成电波的阴影。接收机在移动到阴影区时，电波会发生非常大的阴影衰落，造成通信中断，所以要采取衰落储备，例如：增加发射功率、增加发射天线或减小无线小区的覆盖尺寸等。

2.2.3 对流层散射信道的特点

地球表面覆盖着一层稠密的大气，根据最新 ITU 的定义，大气层分为五层，由地面向上分别为：对流层、平流层（同温层）、中间层、电离层和外层。对流层距离地球表面至大约 10 km（8~17 km）的高度，冷暖空气团对流，有风雷雨电等气象现象，如图 2-7 所示。

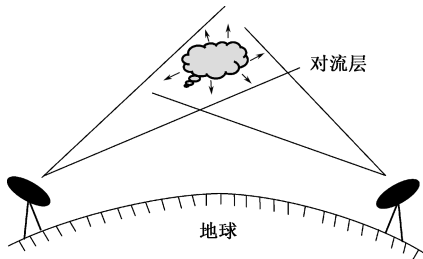


图 2-7 对流层散射传播示意图

几十年来，对于对流层散射传播的机理尚无一种统一的解释，不过散射通信专家们根据大量的电波传播试验数据，早已提出了 3 种不同的假设：散射理论、反射理论和多模理论，这 3 种理论都有大量实验数据可以说明，不过支持散射理论的数据更充分。

散射论者认为，在对流层中不断产生大气涡流（湍流），使温度、湿度和气压产生随机变化，引起大气折射指数发生变化。这种折射指数不断起伏的区域构成散射体。散射体内的大量有极分子，它们受到发射电波的激励后相当于一个个小偶极子天线，按照偶极子天线的方向图产生二次辐射。这些偶极子天线的排列是杂乱无章的，随机变化的。只有其中少量“偶极子天线”的轴向与发射天线的线极化方向一致才能受到最大的激励，它们的二次辐射波最强，形成前向散射波；也只有这少量“偶极子天线”的轴向与接收天线的线极化方向一致，因而其落到接收天线的主波束内的二次辐射波才能被接收天线有效接收，从而形成超视距传播。

散射波传播主要特性：

- （1）对流层散射传播不受核爆炸以及太阳黑子、电离层扰动、极光、磁暴等自然现象的影响，在军事通信中不可或缺。
- （2）对流层散射主要用在 100 MHz~10 GHz 的频段，传播距离 300~800 km，适用于无法建立微波中继站的地区（如：海岛之间，跨越湖泊、沙漠、雪山的地区）。
- （3）通信容量、质量比短波通信高，早在 1995 年已见报导，Macorni 公司（英）生产的 H7600 已达 16 Mb/s，能传数码化电视信号；Aydin（美）生产的 MTDS 达 6.3/12.6（Mb/s）。
- （4）对流层散射信道中电波传播的特点和散射通信为窄波束定向通信决定了这种通信方式的保密性、抗毁性和抗干扰性能均比卫星通信好。
- （5）对流层散射通信的缺点：传播损耗较大，存在多径传输和衰落。

2.2.4 电离层折射信道的特点

大气分子由于受太阳紫外线和高能粒子辐射，而产生游离，分裂为带电电子、离子，由这些带电电子、离子组成的大气层称为电离层，通常距地面 50~500 km。

另外，地球周围的大气层并不是一种均匀媒质，大气的密度、压力、温度与湿度都是随高度而变化的。电波在大气层中的传播，实际上是一个电波在不均匀媒质中的传播问题，电波在传播过程中会发生大气折射现象。

利用电离层折射通信的电波传播方式，如图 2-8 所示，将电离层分成许多厚度极薄的平

行薄层，传播满足几何光学近似条件，电波的折射角度、强度与入射波的角度、波长、强度以及电离层的高度和浓度等因素有关，首先可以用折射理论分析，折射角越来越大，当电波入射角 $\beta$ 小，则临界最高被折射频率 $f_{\max}$ 低，否则，电波会穿过整个电离层，形成透射。折射通信时希望电离层透射出的能力要小，卫星通信时希望电离层透射出的能力要大。

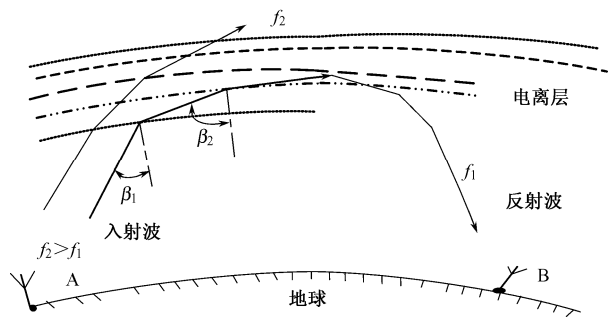


图 2-8 电离层折射通信的电波传播方式

其次，电离层的高度和浓度，随地区、季节、时间、太阳黑子活动等因素的变化而变化，也对通信的临界频率产生影响，电离层的浓度高时折射的频率高，浓度低时折射的频率低，具体见表 2-3 所示。

表 2-3 电离层各层主要数据

|                            | D                   | E                                    | F1                                     | F2                                   |
|----------------------------|---------------------|--------------------------------------|--|--------------------------------------|
| 夏季白天高度/km                  | 60~90               | 85~150                               | 150~200                                | >200                                 |
| 夏季夜间高度/km                  | 消失                  | 90~140                               | 消失                                     | >150                                 |
| 冬季白天高度/km                  | 60~90               | 85~150                               | 160~180（常消失）                           | >170                                 |
| 冬季夜间高度/km                  | 消失                  | 90~140                               | 消失                                     | >150                                 |
| 白天最大电子浓度/个/cm <sup>3</sup> | 约 2×10 <sup>3</sup> | 5×10 <sup>4</sup> ~2×10 <sup>5</sup> | 2×10 <sup>5</sup> ~4.5×10 <sup>5</sup> | 8×10 <sup>5</sup> ~2×10 <sup>6</sup> |
| 夜间最大电子浓度/个/cm <sup>3</sup> | 消失                  | 10 <sup>3</sup> ~10 <sup>4</sup>     | 消失                                     | 10 <sup>5</sup> ~3×10 <sup>5</sup>   |
| 电子浓度最大值处的高度/km             | 约 90                | 115                                  | 约 180                                  | 200~350                              |
| 白天临界频率/MHz                 | <0.4                | <3.6                                 | <5.6                                   | <12.7                                |
| 夜间临界频率/MHz                 |                     | <0.6                                 |  | <5.5                                 |

白天电子密度高，晚上电子密度低，D 层夜晚消失，E 层相对稳定，F1 层夜晚与 F2 合并，F2 层变化复杂。图 2-9 所示为一月份不同纬度处 F2 层 24 小时电子浓度的高度。

综合前述，可以得出 F2 层折射时最高可用频率与通信距离和时间关系，见图 2-10 所示，结合图 2-8 看出，频率过低的电磁波会受到很大的吸收损耗，频率过高的电磁波将穿越电离层不再回到地面，频率为 1.5~30 MHz 的中波和短波都能借助电离层的折射传播较远的距离。

总之，影响电离层折射传播媒质的主要因素有：地球上不同纬度、太阳黑子、核爆炸、季节和气候的变化等。

在利用电离层折射通信时，通信距离越近，要求电波的入射角越小，其最高可用频率将越低。

一般把采用一个折射区进行通信，可以近似认为是一次反射通信，或称为一跳通信，当通信距离较远时，将会有 2 到 3 个折射区，近似认为二次、三次反射通信，或称为两跳、三跳。

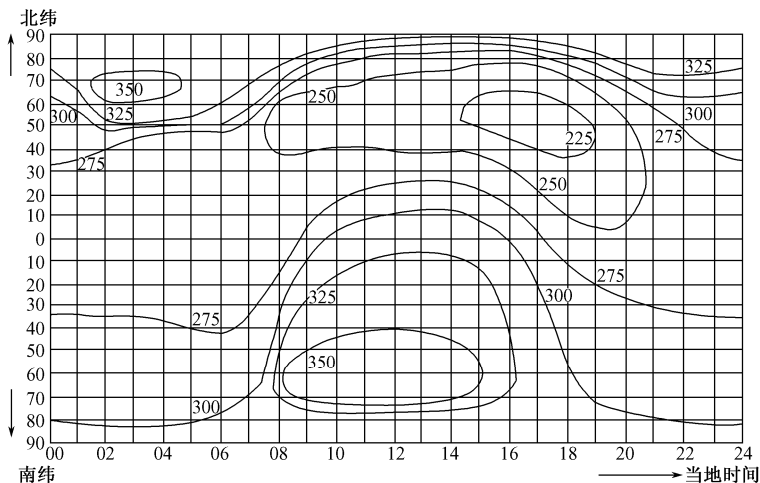


图 2-9 一月份 F2 层高度  $H$  的近似值

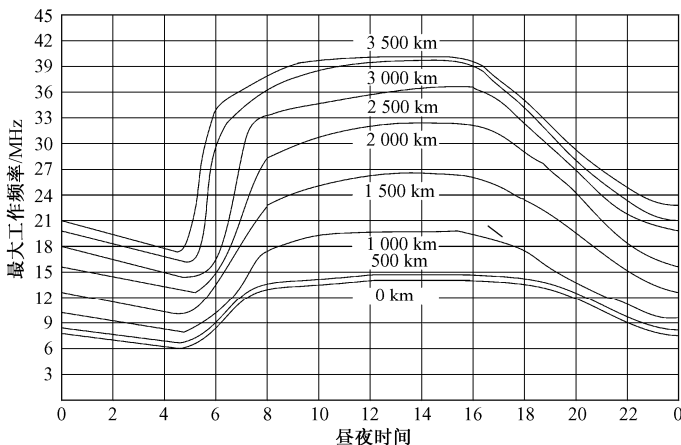


图 2-10 最高可用频率与通信距离和时间关系

## 2.2.5 卫星通信信道的特点

卫星通信是指利用人造地球卫星作为中继站转发无线电信号，根据卫星距离地面的高度分为超过 20 000 km 以上高轨道卫星系统、5 000~20 000 km 的中轨道卫星系统和 500~5 000 km 的低轨道卫星，即卫星通信实现了地面与电离层以外空间的电波传输，如图 2-11 所示，其信道特点一定不同与其他无线信道。

首先，卫星通信的电波要穿透大气层，进入外层空间，本身通信距离长，信号在传输过程中损耗和延时就大。

其次，电波将受到电离层中自由电子、离子以及对流层中的氧分子、水蒸气分子和云、雾、雨、雪等的吸收而产生损耗，衰减在  $10^{20}$  倍以上，故对电波的穿透能力提出较高要求，经过测量，在 0.3~10 GHz 频段，大气损耗最小，透射出的能力最大，故称此频段为“无线电窗口”，在 30 GHz 附近也有一个损耗低谷，通常称此频段为“半透明无线电窗口”。



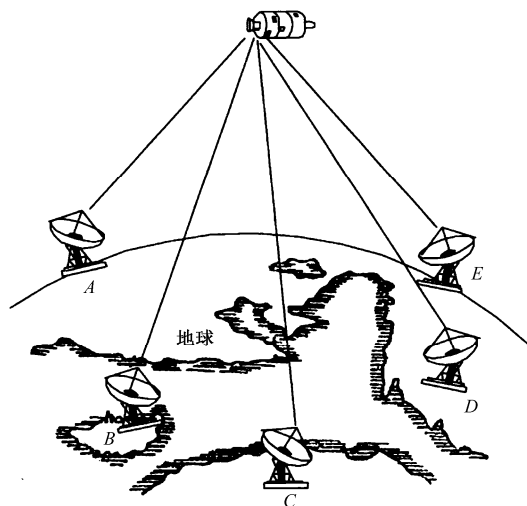


图 2-11 卫星通信示意图

一般工作频段选择在  $1\sim 10\text{ GHz}$ ；最理想的频率在  $4\sim 6\text{ GHz}$ 。

另外，宇宙噪声、太阳黑子、核爆炸等因素对卫星通信影响大。卫星与地球的相对运动也会产生多普勒效应，静止卫星除外。

静止卫星存在星蚀和日凌中断等问题。在每年的春分及秋分前后的各 20 多天中，当同步卫星的星下点（卫星与地心的连线同地球表面的交点）进入当地时间午夜前后，此时卫星、地球和太阳共处一条直线上，地球挡住了阳光，卫星进入地球的阴影区，这种现象称为星蚀。每年春分和秋分前后若干天内，同步卫星的星下点进入当地中午前后的一段时间里，卫星处于太阳和地球之间，地面站天线对准卫星的同时，可能也会对准太阳，这时强大的太阳噪声会使通信无法进行，这种现象称为日凌中断。

## 2.3 广义信道的数学模型

### 2.3.1 狭义信道与广义信道

#### 1. 狭义信道

通常，我们将接在发送端设备和接收端设备中间的传输媒质称为狭义信道，狭义信道定义直观，易理解，如图 2-12 所示。

按照媒质特性的不同，狭义信道通常可分为导向传输信道和非导向传输信道。

导向传输信道有光纤光缆、电缆、双绞线等，信号被束缚在固体介质内传播。

非导向传输信道，包括无线电波、红外线、可见光波等，在空间或宇宙中自由直线方式传播，传播速度为  $3\times 10^8\text{ m/s}$ ，同时具有折射、反射、绕射和干涉等波的特性。

#### 2. 广义信道

在通信理论的分析中，从研究消息传输的观点看，我们所关心的只是通信系统中的信号

传输的结果问题，因而，信道的范围还可以扩大。它除包括传输媒质外，还可能包括有关的转换器，如馈线、天线、调制解调器等等。通常将这种扩大了范围的信道称为广义信道，如图 2-12 所示。

广义信道是从研究各种实际通信问题出发而构成的，根据研究对象和关心问题的不同，可以定义各种形式的广义信道，例如调制信道、数字编码信道和数据链路信道。

### 3. 调制信道

调制信道是从研究调制和解调的基本问题出发而构成的，它的范围是从调制器输出端到解调器的所有转换器及传输媒质，不管其中间过程如何，它们不过是把已调信号进行某种变换而已，我们只需关心变换的最终结果，而无须关心形成这个最终结果的详细过程。因此，研究调制与解调问题时，定义一个调制信道是方便和恰当的，如图 2-12 所示。

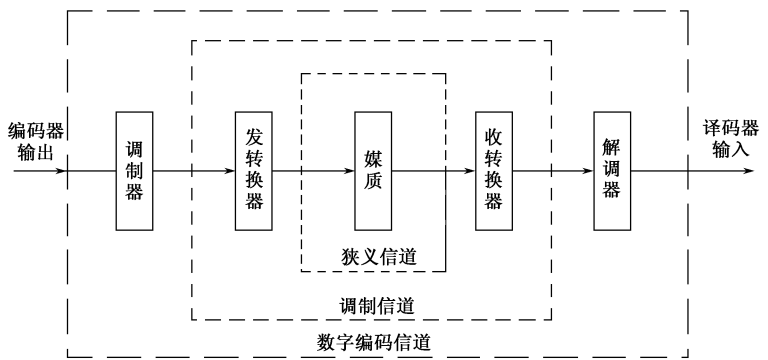


图 2-12 狭义信道、调制信道与数字编码信道

### 4. 数字编码信道

在数字通信系统中，如果仅着眼于数字编码和译码问题，则可得到另一种广义信道——数字编码信道。这是因为，从编码和译码的角度看。编码器的输出是某一数字序列，而译码器输入同样也是一数字序列，它们在一般情况下是相同的数字序列。因此，从编码器输出端到译码器输入端的所有转换器及传输媒质可用一个完成数字序列变换的方框加以概括，此方框称为编码信道，如图 2-12 所示。

### 5. 数据电路信道

在数据通信中，依据数据通信协议，利用数据传输技术，通过数据电路将分布在远端的数据终端设备连接起来，实现数据信号传输、交换、存储和处理。

如果仅着眼于数据传输结果，则可以将收、发端的数据电路终端设备之间的所有器件及传输媒质，叫数据电路信道，如图 2-13 所示。

这是因为，从数据的传输角度看，发端的数据电路终接设备处理的是某一数据序列，而收端的数据传输控制器处理的同样也是一数据序列，它们在一般情况下是相同的数据序列。因此，从发端到收端的数据电路终接设备之间的所有器件及传输媒质可用一个完成数据电路方框加以概括，此方框就可称为数据电路信道。

常用的电话线路调制解调器（Modem）就是一种数据电路终端设备，它是调制器和解调

器的结合，并实现电路的数据传输。发送时，调制器把数据信号转换成适合于电话线路上传输的模拟信号；接收时，模拟信号由解调器将它还原成数据信号，并送到 DTE。当数据信号在数字线路上传输时，DCE 的位置上不再需要 Modem，而改为数据服务单元（Data Service Unit, DSU）。DSU 的主要功能是信号码型与电平的转换、信道特性的均衡、收发时钟的形成与供给，以及线路接续控制等。

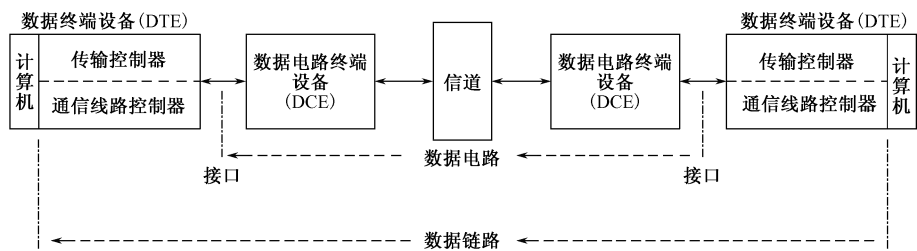


图 2-13 数据电路与数据链路信道

## 6. 数据链路信道

数据电路加上数据传输控制器的功能后就构成了数据链路信道，无线信道加上数据传输控制器的功能称为无线数据链路信道，如图 2-13 所示。

对于数据终端设备来说，数据链路信道着眼于数据传输控制，发端的数据传输控制器输出的是某一数据序列，而收端的数据传输控制器输入的同样也是一数据序列，它们在一般情况下是相同的数据序列。因此，从发端到收端的数据传输控制器之间的所有器件及传输媒质可用一个完成数据序列方框加以概括，此方框就可称为数据链路信道。

### 2.3.2 调制信道的数学模型

通常，为了方便通信系统研究，信道特性描述，引入信道的数学模型。

在频带传输系统中，已调信号离开调制器便进入调制信道。

对于调制和解调而言，通常可以不管选用了什么样的传输媒质、什么样的转换器，以及发生了怎样的传输过程，我们仅关心已调信号通过调制信道后的最终结果。因此，把调制信道概括成一个模型是可能的。

通过对调制信道进行大量的考察之后，可发现它有如下主要特性：

(1) 有一对（或多对）输入端，则必然有一对（或多对）输出端。最基本的是端对端的单输入单输出信道（SISO），最一般的是多输入多输出信道（MIMO）。

(2) 绝大部分信道是线性的，即满足叠加原理。从另一个角度说，线性或近似线性是我们对信道设计的要求；当然，不同的调制方式对信道线性的要求是不同的。

(3) 信号通过信道需要一定的迟延时间，通信距离越远，信号迟延时间越长。

(4) 即使没有信号输入，在信道输出端也可能有一定的功率输出——噪声或干扰，信道噪声有宇宙噪声、工业噪声、线路热噪声等，信道干扰有工业干扰、人为干扰等多种。

一般通信系统分析时主要考虑噪声的影响，现在随着工业干扰、人为干扰的增加，电磁环境的复杂，也要考虑干扰的抑制与抵消。

(5) 信道对信号有损耗（落），距离越长，损耗越大，并且信道的损耗会随时间、环境

等因素的不同而变化。

由此看来，可用一个二对端（或多对端）的时变线性网络去替代调制信道。这个网络就称为调制信道模型，如图 2-14 所示。

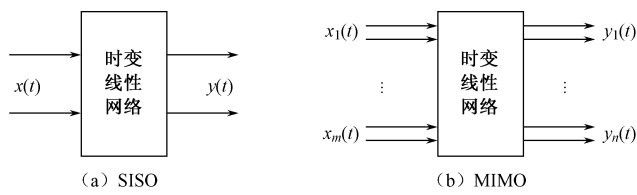


图 2-14 调制信道数学模型

对于二对端的信道模型来说，输入的已调信号  $x(t)$ ，信道输出信号  $y(t)$ ，信道噪声（或信道干扰） $n(t)$ ，则它的输入和输出之间的时域关系式可表示成

$$y(t)=f[x(t)]+n(t) \tag{2-1}$$

式 (2-1) 中， $f[x(t)]$ 表示信道对信号影响（变换）的某种函数关系。

由于  $f[x(t)]$ 形式是个高度概括的结果，非常复杂，为了进一步理解信道对信号的影响，通过简化分析，我们设想它的输入和输出之间的关系式可为

$$y(t)=K(t) \cdot x(t)+ n(t) \tag{2-2}$$

式中， $K(t)$ 表示信道对信号影响（变换）的简化传递函数，它依赖于网络的特性，对信号  $x(t)$ 影响较大。

这样，信道对信号的影响可归纳为两点：一是乘性干扰  $K(t)$ 的影响；二是加性  $n(t)$ 的影响。

如果了解  $K(t)$ 和  $n(t)$ 的特性，则信道对信号的具体影响就能搞清楚，不同特性的信道，仅反映信道模型有不同的  $K(t)$ 及  $n(t)$ 而已。

实际应用中，对于  $n(t)$ 的分析比较简单， $n(t)$ 的主要成分为高斯分布的白噪声，各种信道的  $n(t)$ 差异不大；而  $K(t)$ 则是个很复杂的函数，而且不同信道的  $K(t)$ 差异也很大，它可能包括各种线性畸变、非线性畸变。另外，由于信道的迟延特性和损耗特性随时间作随机变化，所以  $K(t)$ 往往只能用随机过程来表示，因此分析难度较大。但是，这并不意味着是  $K(t)$ 就没有办法分析了，经大量观察发现，有的信道的是  $K(t)$ 不随时间而变，或者随时间变化得极为缓慢，而有的信道的  $K(t)$ 则随时间作随机性的快变化。如果我们将  $K(t)$ 按其随时间变化的快慢不同进行分类，则可将调制信道分为恒参信道与变参信道两类。

恒参信道也称为静态信道，其信道  $K(t)$ 可看成不随时间、环境等因素变化或变化极为缓慢的一类信道。对于理想恒参信道， $K(t)$ 为常数，对应信道幅频特性为常数，相频线性；对于非理想恒参信道， $K(t)$ 变化极为缓慢，对应信道幅频特性会少量畸变，相频特性也会少量畸变。

一般来说，对称型电话电缆、双绞型电缆（计算机网线）、光纤（缆）、波导、中长波地表面传播、超短波及微波视距传播、卫星中继等传输媒质构成的信道是恒参信道。

变参信道，也称随参信道、时变信道，是  $K(t)$ 随时间随机变化的信道。变参信道的特性比恒参信道要复杂得多，对信号的影响也要严重得多。

一般情况下，短波电离层反射、对流层散射、地面移动通信等无线信道都为变参信道。

2.3.3 数字编码信道的数学模型

编码信道是包括调制信道和调制器、解调器在内的信道。它与调制信道模型有明显的不同：即调制信道对信号的影响是通过  $K(t)$ 和  $n(t)$ 使调制信号发生“模拟”变化，而编码信道对信号的影响则是一种数字序列的变换，即把一种数字序列变为另一种数字序列。故有时把编码信道看成是一种数字信道。

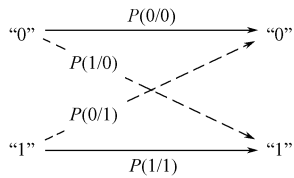


图 2-15 编码信道数学模型

由于编码信道包含调制信道，因而它同样要受到调制信道以及信道噪声的影响。但是，从编/译码的角度看，这些影响已被反映在解调器的最终结果里——使解调器输出数字序列以某种概率发生差错。显然，如果调制信道越差，即特性越不理想，加性噪声越严重，则发生错误的概率将会越大。由此看来，编码信道的模型可用数字信号的转移概率来描述。例如，在常见的二进制数字通信系统中，一个简单的编码信道模型如图 2-15 所示。

这个模型里，把  $P(0/0)$ 、 $P(1/0)$ 、 $P(0/1)$ 、 $P(1/1)$ 称为信道转移概率，具体地把  $P(0/0)$ 和  $P(1/1)$ 称为是正确转移概率，而把  $P(1/0)$ 和  $P(0/1)$ 称为错误转移概率。根据概率性质可知

$$P(0/0)+P(1/0)=1 \tag{2-3}$$

$$P(1/1)+P(0/1)=1 \tag{2-4}$$

转移概率完全由编码信道的特性所决定，一个特定的编码信道就会有相应确定的转移概率。应该指出，编码信道的转移概率一般需要对实际编码信道作大量的统计分析才能得到。编码信道可细分为无记忆编码信道和有记忆编码信道。

无记忆编码信道中，当前码元发生差错与其前后码元的差错没有依赖关系。

有记忆编码信道是指信道中码元发生差错的事件不是独立的，码元发生错误与其前后是有联系的。

2.4 恒参信道及其对所传信号的影响

2.4.1 线性畸变与改善措施

通常，恒参信道的频率特性不理想造成的畸变叫线性畸变，它主要是因为恒参信道的幅度—频率特性和相位—频率特性不理想造成的。线性畸变和非线性畸变的区别是线性畸变不会产生新的频率成分。

现在我们以典型有线电话电缆信道为例，来说明恒参信道的线性畸变。

1. 幅度—频率特性及畸变

幅度—频率畸变是信道的幅度—频率特性的不理想所引起的，这种畸变又称为频率失真。

在通常的电话信道中可能存在各种滤波器，尤其是带通滤波器，还可能存在混合线圈、串联电容器和分路电感等，因此电话信道的幅度—频率特性总是不理想。

例如，图 2-16 为典型音频电话信道的总衰减频率特性。图 2-16 中，低频截止频率约从

300 Hz 开始, 300 Hz 以下每倍频程衰耗升高 15~25 dB; 在 300~1 100 Hz 范围内衰耗比较平坦; 在 1 100~2 900 Hz 内, 衰耗通常是线性上升的; 在 2 900 Hz 以上, 衰耗增加很快, 每倍频程增加 80~90 dB。

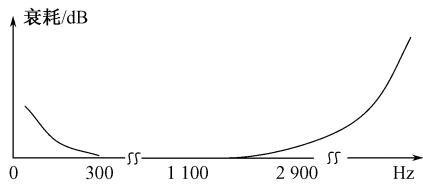


图 2-16 典型音频电话信道的总衰耗频率特性

十分明显, 如上所述的不均匀衰耗必然使传输信号的幅度—频率发生畸变, 引起信号波形的失真。此时若要传输数字信号, 还会引起相邻数字信号波形之间在时间上的相互重叠, 即造成码间串扰。

2. 相位—频率特性及畸变

所谓相位—频率畸变, 是指信道的相位—频率特性偏离线性关系所引起的畸变, 图 2-17 中偏离虚线的部分。

电话信道的相位—频率畸变主要来源于信道中的各种滤波器及感性线圈, 尤其在信道频带的边缘, 相位—频率畸变就更严重。

信道的相位—频率特性还经常采用群延迟—频率特性来衡量。所谓群延迟—频率特性, 它被定义为相位—频率特性的导数, 即

$$\tau(\omega) = \frac{d\varphi(\omega)}{d\omega}$$

(2-5)

我们可以看出, 如果  $\varphi(\omega) \sim \omega$  呈现线性关系 (图 2-6 虚线), 则  $\tau(\omega) \sim \omega$  将是一条水平直线 (图 2-18 虚线)。即信道不发生相位—频率畸变时, 不同频率的信号将有相同的延迟。

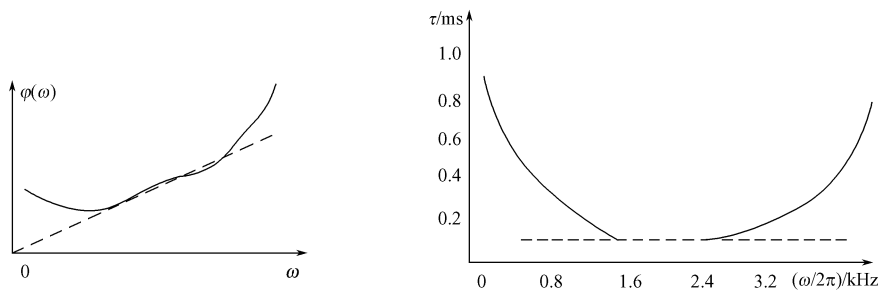


图 2-17 电话信道的相位—频率特性

图 2-18 电话信道的群延迟—频率特性示意图

但实际的信道特性总是偏离理想特性的, 例如一个典型的电话信道的群延迟—频率特性如图 2-18 实线, 即不同频率分量信号将有不同的延迟, 使它们的到达时间造成先后不一, 从而引起信号的畸变。

相位—频率畸变对模拟话音通道影响并不显著, 这是因为人耳对相位变化不太灵敏, 即使存在相移失真也不会感觉到。所以, 在模拟通信系统内往往只注意幅度失真, 而将相移失真放在忽略的地位。但对数字传输却不然, 尤其当传输速率比较高时, 相位—频率畸变将会

引起严重的码间串扰，给通信带来很大损害。

为了减小幅度—频率和相位—频率线性畸变，可采取以下措施：

(1) 要严格限制已调制信号的频谱，使它保持在信道的线性范围内传输。

(2) 在设计总的信道传输特性时，改善信道中的滤波性能，一般都要求把幅度—频率畸变和相位—频率线性畸变控制在一个允许的范围內。

(3) 通过增加一个线性补偿网络，使衰耗特性曲线变得平坦。这一措施通常称之为“均衡”。均衡的方式有频域均衡和时域均衡，具体技术将在本书有关章中介绍。

## 2.4.2 非线性畸变和其他影响

线性畸变是恒参信道对信号的主要影响。除此而外，信道对信号的损害还有非线性畸变、频率偏移及相位抖动、回波干扰等。

非线性畸变主要由信道中的元器件（如磁芯，电子器件等）的非线性特性引起，造成失真或产生寄生频率等。

频率偏移通常是由于通信系统中接收端解调器载波与发送端调制载波之间的频率有偏差（例如，解调载波可能没有锁定在调制载波上），而造成信道传输的信号的每一分量可能产生的频率变化。

相位抖动也是由调制和解调载波发生器的不稳定性造成的，这种抖动的结果相当于发送信号附加上一个小指数的调频。

回波干扰是由于混合线圈不平衡所致。

以上的非线性畸变一旦产生，一般均难以排除。这就需要在进行系统设计时从技术上加以重视。

## 2.4.3 噪声干扰

恒参信道的传输函数  $K(\omega)$  不随时间而变，或变化极为缓慢，所以噪声干扰主要是加性噪声  $n(t)$ 。

加性噪声  $n(t)$  分为高斯白噪声和脉冲噪声两类。高斯白噪声是指噪声的概率密度函数服从高斯分布，功率谱服从均匀分布的一类噪声。通信系统中的传输媒质热噪声、设备热噪声和散弹噪声以及由接收天线收到的辐射等均为高斯白噪声，它是较稳定的、不可避免的背景噪声。

脉冲型噪声是指重复出现的持续时间甚短的脉冲波形，可以是人为产生，也可以是自然产生，大部分脉冲噪声是因电气切换和通信设备开关的瞬态引起的。另外，微波衰落和维护工作中的偶然碰撞也会产生较长的突发脉冲噪声；天然雷电也会引起脉冲噪声；这些噪声均为随机性影响，无法彻底消除，只能增加抗干扰能力来减轻它的影响。

## 2.5 变参信道及其对所传信号的影响

卫星通信信道、短波电离层反射、对流层散射、地面移动通信等无线信道都为变参信道。变参信道的特性比恒参信道要复杂得多，对信号的影响也要严重得多。其根本原因在于它包

含一个非常复杂的传输媒质。

变参信道传输媒质通常具有以下特点：

- (1) 存在多径传播现象，造成多径衰落；
- (2) 当发送和接收终端相对块速移动时，产生多普勒效应；
- (3) 传输特性随环境与位置的不同也不同。

2.5.1 多径传播、频率弥散和选择性衰落

卫星通信、微波移动通信、电离层反射和散射、对流层散射等无线通信时，始终存在着可能不断变化的反射和散射物体，直射波、反射波、散射波会经多条路径传播到接收端，这种现象称多径传播，如图 2-19 所示。由于到达接收天线的信号幅度、相位、时延以及入射角度不同，造成接收到的信号合成后起伏很大，使信号产生深度且快速的衰落，这就是多径衰落。多径衰落包络分布规律服从瑞利分布。

设发射信号为等幅正弦波  $A\cos\omega_0t$ ，经过  $n$  条路径传播后的接收信号  $R(t)$ 可表示为

$$\begin{aligned} R(t) &= \sum_{i=1}^n a_i(t) \cos \omega_0[t - t_{di}(t)] \\ &= \sum_{i=1}^n a_i(t) \cos[\omega_0t + \varphi_i(t)] \end{aligned} \tag{2-6}$$

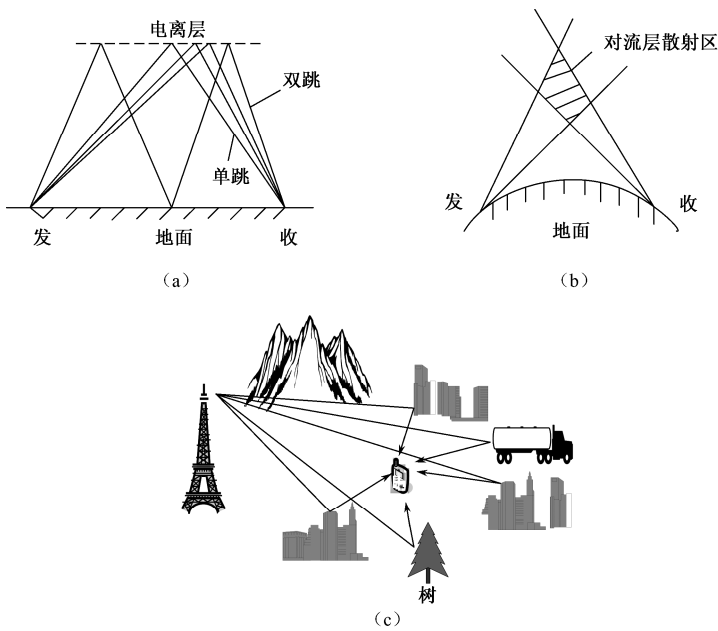


图 2-19 多径传播示意图

式中， $a_i(t)$ ——第  $i$  条路径到达接收端的随机幅度； $t_{di}(t)$ ——第  $i$  条路径对应的延迟时间； $\varphi_i(t)$ ——相应的随机相位，它等于  $\varphi_i(t) = -\omega_0t_{di}(t)$ 。

$a_i(t)$ 和 $\varphi_i(t)$ 随  $t$  的变化要比 $\omega_0$ 的周期变化慢得多，因此式 (2-6) 又可写成



$$\begin{aligned}
 R(t) &= \sum_{i=1}^n a_i(t) \cos \omega_0 [t - t_{di}(t)] = \sum_{i=1}^n a_i(t) \cos [\omega_0 t + \varphi_i(t)] \\
 &= [\sum_{i=1}^n a_i(t) \cos \varphi_i(t)] \cos \omega_0 t - [\sum_{i=1}^n a_i(t) \sin \varphi_i(t)] \sin \omega_0 t \\
 &= a_I(t) \cos \omega_0 t - a_Q(t) \sin \omega_0 t \\
 &= a(t) \cos [\omega_0 t + \varphi(t)]
 \end{aligned} \tag{2-7}$$

其中  $a(t) = \sqrt{a_I^2(t) + a_Q^2(t)}$ ,  $\varphi(t) = \tan^{-1} \left[ \frac{a_Q(t)}{a_I(t)} \right]$ 。

由式 (2-7) 可知:

- 等幅信号  $\rightarrow$  调幅信号;
- 单一频率  $\rightarrow$  许多频率分量 (频率弥散、色散);
- $R(t)$  具有选择性衰落 (下面证明)。

设发送的信号为  $f(t) \leftrightarrow F(\omega)$

假定两条传输路径, 收端两信号幅度相等, 只存在时延  $\tau$ , 则

第一条传输路径  $Kf(t - t_0) \leftrightarrow KF(\omega)e^{-j\omega t_0}$

第二条传输路径  $Kf(t - t_0 - \tau) \leftrightarrow KF(\omega)e^{-j\omega(t_0 + \tau)}$

两条路径传播时的接收信号为  $R(t) = K[f(t - t_0) + f(t - t_0 - \tau)]$

$$R(t) \leftrightarrow KF(\omega)e^{-j\omega t_0} [1 + e^{-j\omega\tau}]$$

$$H(\omega) = \frac{R(\omega)}{F(\omega)} = Ke^{-j\omega t_0} [1 + e^{-j\omega\tau}]$$

$$\begin{aligned}
 |H(\omega)| &= |Ke^{-j\omega t_0} (1 + e^{-j\omega\tau})| = K |1 + e^{-j\omega\tau}| \\
 &= 2K \left| \cos \frac{\omega\tau}{2} \right|
 \end{aligned} \tag{2-8}$$

依式 (2-8) 画出两条路径传播时选择性衰落特性, 如图 2-20 所示。

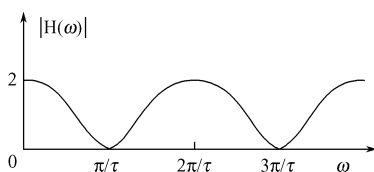


图 2-20 两条路径传播时选择性衰落特性

通过上述分析和图 2-20 可以得出如下结论:

- (1) 多径传播使单一载频信号变成了包络和相位都变化 (实际上受到调制) 的窄带信号, 也就是说, 多径传播引起了频率弥散, 使单个频率变成了一个窄带频谱。
- (2) 多径传播会引起频率选择性衰落。所谓频率选择性衰落是指信号频谱中某一些分量衰减特别大 (传输增益为 0), 而另一些频谱分量衰减却比较小, 经过传输后信号出现的畸变。
- (3) 多径传播会引起时间选择性衰落。所谓时间选择性衰落是指时延  $\tau$  不同, 衰落不同。
- (4) 多径传播会引起空间选择性衰落。所谓空间选择性衰落是指空间环境的不同, 衰落不同。

由于信号在无线信道传播时，其所经历的衰落取决于发送信号和信道的特性，也就是信号参数与信道参数决定了不同的发送信号将经历不同类型的衰落。所以在不同的空间、不同的频率和不同的时间其衰落特性是不一样的。

根据发送信号的周期（频率）与信道变化快慢程度的比较，可以将信道分为快衰落信道和慢衰落信道。快衰落信道的冲激响应在信号周期内变化很快，也就是信道的相干时间比发送信号的周期短。慢衰落信道的冲激响应变化率大大小于发送的基带信号的周期，因此，可以认为在一个或几个信号周期内，信道为非时变的。

通常，多径效应引起的选择性衰落为快衰落，电离层浓度变化等因素所引起的信号衰落为慢衰落，亦称为平坦性衰落。

### 2.5.2 多普勒频移与多普勒衰落

当运动的物体达到一定的速度时（如急速行使的汽车、超音速飞机、人造卫星在发射等），固定点接收到的从运动体发来的载波频率将随物体运动速度的不同产生不同的频移，导致信号频率扩展，通常把这种现象称为多普勒频移（或扩散、扩展）。

例如，人造卫星在发射前，其星上发射机的载频 $f_0$ 是预知的。发射后，地面接收站收到的载波信号频率已不是 $f_0$ 了，而是 $f_0 + f_d$ 。

根据电磁学的基本理论得知，由于多普勒扩展与相干时间成反比，而相干时间是信道随时间变化的度量，所以多普勒扩展引起的衰落与时间有关，是时间选择性衰落。

只要运动，多普勒频移将一定存在，导致接收端产生附加的随机调频噪声。当运动速度较慢时，多普勒频移较小，影响不大，可忽略不计。当运动速度快时，必须使用采用了“锁相技术”的接收机才行。

### 2.5.3 阴影效应与衰落

当电波在传播路径上遇到建筑物、树木、森林等的阻挡时，则会形成电磁场的阴影。

阴影效应主要表现在树木和建筑物的吸收、散射或绕射引起直射波的衰减变化，以及相关多径分量对直射波的干涉作用，衰减量决定于树叶和枝干的浓密度、电波穿越树冠的路径以及建筑物的大小。

当移动用户通过不同障碍物的阴影时，就造成接收场强中值的变化。这种由于阴影效应导致接收场强中值随着地理位置改变而出现的缓慢变化称为阴影衰落。这种衰落是一种慢衰落，衰落率与移动物的速度以及阻挡物的分布有关。

### 2.5.4 变参信道特性的改善

变参信道特性的改善措施可以发送端或接收端进行，或者在两者共同进行。前者作为预防，后者作为补救，具体方案有冗余法与均衡法。

#### 1. 冗余法

冗余法是指采用一个以上的链路来传输信号，以减少传输失败的概率。有分集技术、纠

错编码、重发技术等。例如：发送分集是同一个信号由若干个天线向同一个接收机发送，接收分集是若干接收天线接收同一个信号。

## 2. 均衡法

均衡法是指通过频域均衡、时域均衡、信源的适应等技术改善信道特性。

两种方法互不排斥，在实际使用时可以组合。

对于平坦性衰落（慢衰落），主要采用加大发射功率和在接收机内采用自动增益控制技术和方法即可。

对于快衰落，通常可采用多种措施，例如，各种抗衰落的调制/解调技术及接收技术等。其中较为有效且常用的抗衰落措施乃是分集接收技术。

## 3. 分集接收技术

按广义信道的含义说，分集接收可看做是变参量信道中的一个组成部分或一种改造形式，而改造后的变参量信道，衰落特性将能够得到明显改善。

下面简单介绍分集接收的原理。

前面说过，快衰落信道中接收的信号是到达接收机的各路径分量的合成。如果在接收端同时获得几个不同的信号，将这些信号采取适当技术合并后使选择性衰落的影响大大减小，这就是分集接收的基本思想。分集两字就是分散得到几个信号并集中（合并）这些信号的意思。只要被分集的几个信号之间是独立的，那么经适当合并后就使系统性能大为改善。

互相独立或基本独立的一些信号，一般可利用不同路径或不同频率、不同角度、不同极化等接收手段来获取，于是大致有如下几种分集方式。

（1）空间位置分集（多天线）。在接收端架设几副天线，天线的相对位置都要求有足够的间距（一般在 100 个信号波长左右），以保证各天线上获得的信号彼此基本独立。

（2）空间角度分集（智能天线）。这是利用天线波束指向不同方向上的信号又不同相关性的原理形成的一种分集方法，例如在微波天线上设置若干个反射器，产生相关很小的几个波束。

（3）频率分集。用多个不同频率传送同一个信息，如果各载频的频差相隔比较远，则各分散信号彼此也基本不相关。

（4）时间分集。对于随机衰落信号，如果对其振幅进行顺序取样，那么在时间间隔大于相干时间的两个采样点是互不相关的。时间分集就是根据这一特点进行的。将信号按间隔一定的时隙重复传输  $L$  次，只要时间间隔大于相干时间，就得到  $L$  条独立的分集支路。

（5）极化分集。这是分别接收两种不同极化（水平极化和垂直极化）而构成的一种分集方法，一般这两种波在信道中相关性极小。

（6）场分集。场分集利用电场和磁场来分别传输信号，以取得互不相关的信号副本。这是由于任一点的电场和磁场分量是互不相关的。

当然，还有其他分集方法，这里就不再叙述。

需要指出的是，分集方法均不是互相排斥的，在实际使用时可以是组合式的。例如，由二重空间分集和二重频率分集组成四重分集系统等。

从总的分集效果来说，分集接收除能提高接收信号的电平外，主要是改善了衰落特性，使信道的衰落平滑了、减少了。例如，无分集时，若误码率为  $10^{-2}$ ，则在用四重分集时，误码率可降低至  $10^{-7}$  左右。由此可见，用分集接收方法对随参信道进行改善是非常有效和必要的。

## 2.6 信道带宽、系统带宽与信号带宽

### 2.6.1 带宽的概念与种类

带宽 (bandwidth)，即通频带宽度，指波长、频率或能量带的范围。

所有带宽，它们均用符号  $B$  表示，单位为  $\text{Hz}$ ，计算方法类似，而表示的概念不同，所以用到带宽时都需要说明是哪种带宽。

对通信系统信号传输的分析中，经常上遇到信道带宽、系统带宽、信号带宽；实际带宽与理想带宽；绝对带宽与相对带宽；窄带与宽带等不同种类概念的带宽。

下面分别说明这些带宽的概念。

(1) 信号带宽：表示信号能量谱密度或功率谱密度在频域的分布规律，由信号的特点决定。

(2) 系统带宽：表示电路系统的频率传输特性，其中在系统带宽内，信号能较好通过；系统带宽外，信号被抑制，系统带宽特性主要由构成系统的电路、电子器件等决定。

(3) 信道带宽：任何实际的信道所能传输的信号频率有一定的范围，这一范围称为该信道频带的宽度，即信道带宽。

(4) 信号带宽、系统带宽与信道带宽之间的大小关系。在模拟信号通信过程中，要求信号不失真传输，一般要求实际带宽的大小关系满足：

$$\text{信号带宽} \leq \text{系统带宽} \leq \text{信道带宽} \tag{2-9}$$

在数字信号通信过程中，不考虑信号不失真问题，仅考虑无码间串扰，要求实际带宽的大小关系满足：

$$\text{信号带宽} \leq \text{系统带宽}/2 \leq \text{信道带宽} \tag{2-10}$$

(5) 实际带宽与理想带宽。在信号与系统理论分析中，信号和系统的频率取值域为  $(-\infty \sim +\infty)$ ，但负频率实际不存在，所以，把实际存在的正频率区域带宽，称为实际带宽。例如，信号实际带宽、系统实际带宽、信道实际带宽等。

理想带宽是指在通信系统分析中，假设信道或系统分析对象为理想情况下的呈现的频带宽度。例，假设通信系统为理想低通滤波器时，对外呈现出的带宽就理想带宽，这样可简化系统的数学理论分析。

图 2-21 所示为系统实际带宽与系统理想带宽示意图。

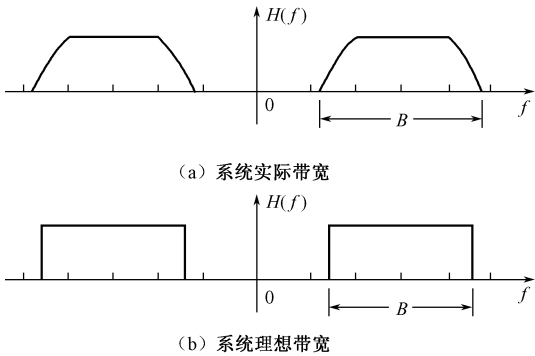


图 2-21 系统实际带宽与系统理想带宽示意图

(6) 绝对带宽与相对带宽。下面以系统带宽为分析对象，说明绝对带宽与相对带宽的概念。

设通信系统的传递函数为  $H(f)$ ，在中心频率为  $f_0$  的  $\Delta f$  宽度内的信号可以完全通过本系统，则称  $\Delta f$  为系统的绝对带宽， $\Delta f/f_0$  为相对带宽，如图 2-22 所示。

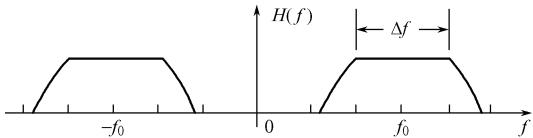


图 2-22 绝对带宽与相对带宽示意图

(7) 窄带、宽带与超宽带。依据带宽的大小可将其分为窄带、宽带与超宽带，但窄带、宽带与超宽带之间没有严格界限，不同通信系统和国家的规定不同，下面给出一个常见的数值，仅供大家参考。

- 窄带： $\Delta f/f_0 < 1\%$ ；
- 宽带： $1\% < \Delta f/f_0 < 25\%$ ；
- 超宽带： $\Delta f/f_0 > 25\%$ （美国的定义）；
- 超宽带： $\Delta f/f_0 \sim 100\%$ （俄国人的定义）。

### 2.6.2 带宽的几种计算方法

下面以信号带宽为分析对象，说明带宽的计算方法。

从理论上讲，除了极个别的信号外，信号的频谱分布都是无穷宽的。例如，矩形脉冲信号  $x(t)$  的频谱函数

$$X(\omega) = F[x(t)] = \frac{\tau \sin(\omega\tau/2)}{\omega\tau/2} = \tau Sa(\omega\tau/2) \tag{2-11}$$

如图 2-23 所示，其频谱是很宽的，可以说是无穷宽的。

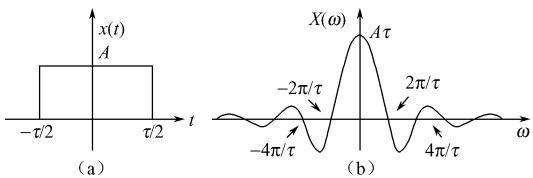


图 2-23 形脉冲信号的频谱

矩形脉冲序列  $x(t)$  的频谱函数  $X(\omega)$ ，见图 2-24，是由许多离散的频率成分组成，而包络和矩形脉冲的频谱函数相同，频谱也是无穷宽的。

如果把凡是有信号频谱的范围都算带宽，那么很多信号的带宽变为无穷大了，显然这样定义带宽是不恰当的，一般信号虽然频谱很宽，但绝大部分实用信号的主要能量（功率）都是集中在某一个不太宽的频率范围以内，因此通常根据信号能量（功率）集中的情况，恰当地定义信号的带宽。这样就要必须知道信号能量（功率）和频率之间的关系。

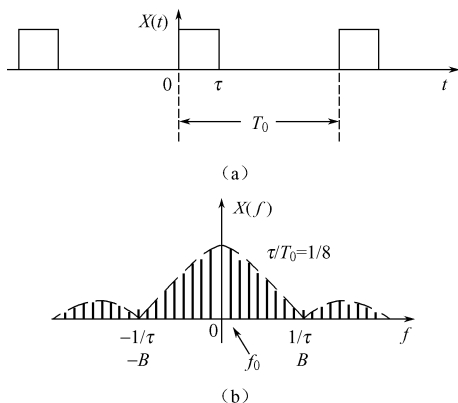


图 2-24 矩形脉冲序列及频谱

根据帕塞瓦尔定理：信号时域总能量等于信号频域各个频率分量单独贡献出来的能量的积分（或信号时域的平均功率等于信号频域各个频率分量单独贡献出来的平均功率的和），这就像能量守恒定理。即对信号  $x(t)$  的能量  $E$  有

$$E = \int_{-\infty}^{\infty} x^2(t) dt = \frac{1}{2\pi} \int_{-\infty}^{\infty} |X(\omega)|^2 d\omega \quad (2-12)$$

为了叙述方便，设信号频谱函数有  $X(\omega) = X(2\pi f) = X(f)$ ，能量谱密度函数  $G(\omega) = G(2\pi f) = G(f)$ ，则有

$$\begin{aligned} E &= \frac{1}{2\pi} \int_{-\infty}^{\infty} |X(\omega)|^2 d\omega = \frac{1}{2\pi} 2 \int_0^{\infty} |X(\omega)|^2 d\omega \\ &= 2 \frac{1}{2\pi} \int_0^{\infty} |G(\omega)| d\omega = 2 \int_0^{\infty} |G(\omega)| df \end{aligned}$$

其中  $G(\omega) = |X(\omega)|^2$ ，则

$$\begin{aligned} E &= \frac{1}{2\pi} \int_{-\infty}^{\infty} |X(\omega)|^2 d\omega \\ &= \frac{1}{2\pi} 2 \int_0^{\infty} |X(\omega)|^2 d\omega = \frac{1}{2\pi} 2 \int_0^{\infty} |X(\pi f)|^2 d(2\pi f) \\ &= 2 \int_0^{\infty} |X(f)|^2 df = 2 \int_0^{\infty} G(f)^2 df \end{aligned} \quad (2-13)$$

用这个公式，就可方便的通过信号的频域表示式得出信号的能量。

下面就以信号能量谱  $G(\omega)$  为分析对象，给出实际带宽常用百分比带宽、3 dB 带宽、等效带宽、脉冲信号带宽的计算方法，信号功率谱  $P(\omega)$  为分析对象的带宽定义方法类似。

### 1. 百分比带宽

设信号的总能量为

$$E = \int_{-\infty}^{\infty} G(f) df$$

在频率  $(-B \sim B)$  范围内信号的能量为  $\int_{-B}^B G(f) df$

两者之间的比值为

$$\gamma = \frac{\int_{-B}^B G(f) df}{E} = \frac{2 \int_0^B G(f) df}{E} \quad (2-14)$$

$\gamma$ 百分比可取 90 %、95 %或 99 %等，相对应的  $B$  就称百分比带宽，如图 2-25 所示。信号带宽、系统带宽和信道带宽一般都可利用百分比带宽的概念计算。

## 2. 3 dB 带宽

对于频率轴上具有明显的单峰形状（或者一个明显的主峰）的能量谱（功率谱）密度的信号，从峰值  $f=0$  处，能量谱（功率谱）密度下降到 1/2 处所对应频率间隔，作为信号带宽。由于能量谱（功率谱）的 0.5 相当 3 dB，所以将这种带宽称为 3 dB 带宽，如图 2-26 所示。

信号带宽、系统带宽和信道带宽一般也可用 3 dB 带宽的概念计算。

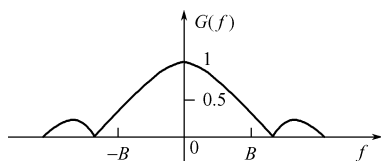


图 2-25 百分比带宽示意图

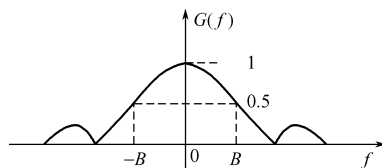


图 2-26 3 dB 带宽示意图

## 3. 等效带宽

假设存在矩形频谱，矩形频谱具有的能量（或功率）与信号的能量（或功率）相等，若以信号频谱  $f=0$  处的高作为矩形频谱的高，则矩形频谱宽的 1/2 对应的频率值，就称为等效带宽，如图 2-27 所示。

信号带宽、系统带宽和信道带宽一般也可用等效带宽的概念计算。

## 4. 脉冲数字信号带宽

在进行脉冲数字信号传输时，还经常用到脉冲主瓣带宽（脉冲带宽），它是指脉冲数字信号的离散频率谱上，第一个 0 点处对应的频率值，如图 2-28 所示。

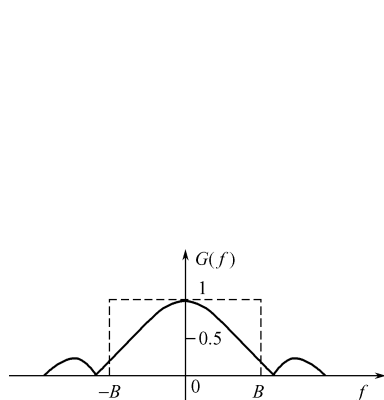


图 2-27 等效带宽示意图

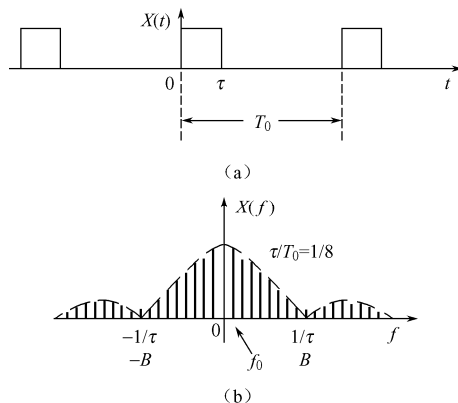


图 2-28 脉冲带宽示意图

一般  $\tau < T_0$ ，则  $B = 1/\tau > 1/T_0 = f_0 = f_b = R_B$ ，即脉冲带宽大于传码率。

当脉冲宽度和脉冲周期相同时， $T_0=\tau$ ，有  $B=1/\tau=1/T_0=f_0=f_b=R_B$ ，即脉冲带宽等于脉冲速率和传码率。

仅数字脉冲信号和数据信号带宽用此概念计算。

## 2.7 信道容量

信道能传输信息的最大能力称为信道容量或最大数据速率。通常以信道每秒所能传输比特数为单位，记为“b/s”、“比特/秒”或“位/秒”。

信道容量越大，信道的传输能力就越强，实际应用中信道容量应大于传输速率，否则高的传输速率得不到充分发挥和利用。

另外，数据信号一般都是依数字信号形式在数字信道或在模拟信道中传输，传输速率受到信道噪声、信道带宽和信道码间串扰等多种因素的影响，故信道容量的分析比较复杂，这里依线性叠加原理，分开考虑各个因素，所得结论可供参考。

### 2.7.1 香农信道容量

在假设信道（或系统）无码间串扰时，当信号与信道加性高斯白噪声的平均功率给定，信道的理想带宽为  $B$ （Hz），理论上单位时间内可能传输的最大信息量  $C$  可以由香农（Shannon）定理确定，香农定理的公式表示式如下：

$$C = B \log_2 \left( 1 + \frac{S}{n_0 B} \right) \quad (\text{b/s}) \tag{2-15}$$

其中信号功率为  $S(\text{W})$ ，加性高斯白噪声单边功率谱密度为  $n_0(\text{W/Hz})$ ，令信道的噪声功率为  $N=n_0B=\sigma^2$ ，则

$$C = B \log_2 \left( 1 + \frac{S}{N} \right) \quad (\text{b/s}) \tag{2-16}$$

结论：

（1）提高  $S/N$  或  $B$ ，则信道容量增加。

（2）给定  $B$ 、 $S/N$  下，信道的极限传输能力为  $C$ ，且此时能做到无差错传输。实际传输速率（一般地）要求不能大于  $C$ ，除非允许存在一定的差错率。

（3）给定  $B$ ， $n_0 \rightarrow 0$ ，则  $C \rightarrow \infty$ ，这意味着不考虑系统码间串扰时，无干扰信道容量为无穷大。

（4） $C$  可以通过  $B$  及  $S/N$  的互换而保持不变。香农定理又告诉我们，维持同样大小的信道容量，可以通过调整信道的  $B$  及  $S/N$  来达到，即信道容量可以通过系统带宽与信噪比的互换而保持不变。

（5）如果  $S$ 、 $n_0$  一定，则无限增大  $B$  并不能使  $C$  值也趋于无限大，这是因为这时的噪声功率  $n_0B$  也趋向于无限大的缘故。

可以证明

$$\lim_{B \rightarrow \infty} C = \frac{S}{n_0} \log_2 e \approx 1.44 \frac{S}{n_0} \quad (\text{b/s}) \tag{2-17}$$

香农定理给出了通信系统所能达到的极限信息传输速率，达到极限信息速率的通信系统



称为理想通信系统。

香农定理描述了无码间串扰时，在有限带宽、有随机热噪声的情况下，信道的最大传输速率与信道带宽、信噪比之间的关系。

由于信道上存在损耗、延迟、噪声，所以会引起信号强度减弱，导致信噪比  $S/N$  降低。延迟会使接收端的信号产生畸变，噪声会破坏信号，产生误码。香农定理指出，对于一定的传输带宽和一定的信噪比，信息传输速率的最大值就确定了。

要想提高信息的传输速率，或者提高传输线路的带宽，或者提高所传信号的信噪比，此外没有其他任何办法。但是，香农定理只证明了理想通信系统的“存在性”，却没有指出这种通信系统的实现方法。

### 2.7.2 奈奎斯特信道容量

奈奎斯特（Nyquist）定理也叫抽样定理，它指的是一个频带限制在  $(0, f_h)$  内的时间连续信号  $x(t)$ ，如果以不大于  $1/(2f_h)$  的间隔，对它进行等间隔抽样，则  $x(t)$  将被所得到的抽样值完全确定。也就是说，如果对上述信号以  $f_s \geq 2f_h$  的速率均匀抽样， $x(t)$  可以被所得后的抽样值完全确定。最小抽样速率  $f_s = 2f_h$  被称为奈奎斯特速率，最大抽样时间间隔  $1/(2f_h)$  被称为奈奎斯特间隔。

奈奎斯特定理用于理想信道，它给出了无噪声时信道带宽对最大数据速率的限制，可表示为

$$C = 2B \log_2 N \quad (\text{b/s}) \quad (2-18)$$

其中  $B$  为理想信道带宽（单位为  $\text{Hz}$ ）， $N$  为数字符号所采取的进制数或基本符号的个数， $C$  为该信道最大的数据速率。

奈奎斯特定理说明，如果对某一带宽有限的时间连续信号进行抽样，且抽样速率达到一定数值时，那么根据这些抽样值就能准确地确定原信号，该定理为模拟信号的数字传输奠定了理论基础。

### 思考与练习

- 2-1 说明狭义信道与广义信道区别。
- 2-2 说明常用传输媒质的工作频率范围与用途。
- 2-3 画出调制信道、编码信道和数据链路信道数学模型，说明它们的区别。
- 2-4 变参信道传输媒质具有哪些特点？
- 2-5 何谓分集接收？分集接收有几种分集方式？
- 2-6 说明信道带宽、系统带宽与信号带宽的区别。
- 2-7 说明带宽的常用计算方法。
- 2-8 说明香农信道容量与奈奎斯特信道容量的区别。

## 第3章 信源与信源压缩编码技术

现代通信系统中信源可以分为3类：计算机产生的数据流、语音产生的数据流和图像产生的数据流，而数据信号的带宽要大于模拟信号的带宽，这将占用较大的信道资源。

为了提高信道（特别是无线信道）的利用率和传输效率，在有限的带宽资源条件下，进行信源压缩编码是最有效的方法，信源压缩可以显著地减少传输所需的带宽，也就可以降低费用和发送时间。

### 3.1 引言

语音、图像和数据等信号都包含大量的冗余，信号的有些部分不包含什么信息，有些可以根据其他部分估算出来，即具有一定的相关性，这种“冗余”是压缩编码的前提。

压缩编码的方法很多，也有不同的分类方法，根据信源特点可分为语音压缩编码、图像压缩编码和数据压缩编码。根据压缩编码的方法可分为：

（1）概率匹配编码。根据信号或消息序列出现概率的分布特性，注重寻找概率与码字长度间的最优匹配，以保证总的代码最短，叫做概率匹配编码或统计编码，或信息熵编码。

（2）变换编码。将信号从一种信号空间变换到另一种更有利于压缩编码的信号空间，然后进行编码。它又可以分为两类：一类是预测变换，利用信号的相关性，预测将来的信号，对预测的误差（余量、残差）进行编码。另一类是函数变换，利用信号在不同函数空间分布的不同，选择合适的函数变换后进行编码。有的文献只将函数变换编码称为变换编码，而将前者称为预测编码。

（3）识别编码。分解文字、语音、图形和图像的基本特征，与汇集这些基本特征的样本集对照识别，选择失真最小的样本编码传送。

实际应用时并不拘泥于哪一类，要根据信号的特点来选择合适的编码方法，往往几种方法并用。

### 3.2 语音波形压缩编码技术

语音压缩编码技术可以分为3类，一是波形压缩编码，二是参量压缩编码，三是混合压缩编码。

波形压缩编码的基本思路是，将时间域信号直接变换为压缩的数字代码，接收端忠实地再现语音的时域波形，为了降低比特率，可充分利用抽样点之间的信息冗余性对差值信号进行编码，在不影响语音质量的前提下，比特率可以降至32 kb/s，其特点是重建的信号质量好，压缩效率低。

参量压缩编码根据对语音形成的机理分析，着眼于构造语音生成模型。该模型以一定精度模拟发出语音的发声过程，发送端对该模型的参量（参数）进行编码，相当于对该语音信号的主要特征而非具体的语音波形幅值编码。接收端，根据该模型还原生成发话者的音素，

根据频域分析可知该模型就对应为具有一定特点的时变线性滤波器。由于模型参数的更新频率较低，并可利用抽样值间的一定相关性，故可有效地降低编码比特率。因此，目前小于 16kb/s 的低比特率语音编码都采用参量编码。它在移动通信、多媒体通信和 IP 网络电话的应用中起到了重要的作用。

混合编码压缩是将波形压缩编码与参量压缩编码的取长补短，完成编码的方案。较为常见的语音压缩编码方案及比特率如表 3-1 所示。

表 3-1 语音编码方案及比特率

|      |                             |  |
|------|-----------------------------|--|
| 波形编码 | 脉冲编码调制（PCM）                 | 码速率：16~64 kb/s，适应能力强、语音质量好。如：程控数字电话交换系统、小灵通、MP3 等。K/TGR146 型超短波电台采用 16 kb/s 或 32 kb/s 的 CVSD |
|      | 自适应差分脉冲编码调制（ADPCM）          |  |
|      | 瞬时压扩 PCM                    |  |
|      | 简单增量调制（ΔM）                  |  |
|      | 连续可变斜率的增量调制（CVSD）           |  |
|      | 子带编码（SBC）                   |  |
| 参量编码 | 共振峰声码器                      | 码速率：0.6/1.2/4.8/9.6 kb/s，语音质量中下。如：多媒体通信、IP 网络电话等   |
|      | 通道声码器                       |  |
|      | 线性预测声码器（LPC）                |  |
|      | 声激励声码器                      |  |
| 混合编码 | 多脉冲激励线性预测编码（MPLPC）          | 码速率：4.8~16 kb/s，语音质量好。如 GSM、CDMA 移动通信网等  |
|      | 规则脉冲激励长时预测线性编码（RPE-LTP-LPC） |  |
|      | 码激励线性预测编码（CELP）             |  |

本节介绍波形压缩编码的代表脉码调制（PCM）及自适应差值脉码调制（ADPCM）。在介绍这些之前，首先介绍抽样、量化与编码的概念，它是一切模拟信号数字化的理论基础。

3.2.1 抽样定理及实际抽样的方法

1. 低通信号的抽样定理

设低通信号  $x(t)$ ，频率分布在  $(0, f_H)$ ，如果以  $f_s \geq 2f_H$  的抽样速率进行均匀抽样，则抽样信号  $x_s(t)$  携带着  $x(t)$  的全部信息，即从  $x_s(t)$  中可恢复出原信号。当  $f_s = 2f_H$  时，称  $f_s$  为奈奎斯特速率。

2. 带通抽样定理

设带通信号  $x(t)$  的上截止频率为  $f_H$ ，下截止频率为  $f_L$ ，如果以下式抽样速率  $f_s$  进行均匀抽样：

$$f_s \geq 2(f_H - f_L)(1 + k/n) = 2B(1 + k/n) \tag{3-1}$$

其中， $B = f_H - f_L$ ， $k = [f_H / (f_H - f_L)] - n$ ， $n$  为不超过  $f_H / (f_H - f_L)$  的最大正整数，则抽样信号  $x_s(t)$  携带着  $x(t)$  的全部信息，即从  $x_s(t)$  中可恢复出原信号。

3. 实际抽样的方法

抽样定理中所说的抽样信号为一个理性（脉冲）冲激信号，实际脉冲总是有一定的宽度，

则根据抽样后脉冲幅度的变换情况可分为：理想抽样、自然抽样与平顶抽样，如图 3-1 所示。其中，自然抽样的脉冲顶部随  $x(t)$  规律变化而变，平顶抽样——脉冲顶部在持续时间内保持不变。

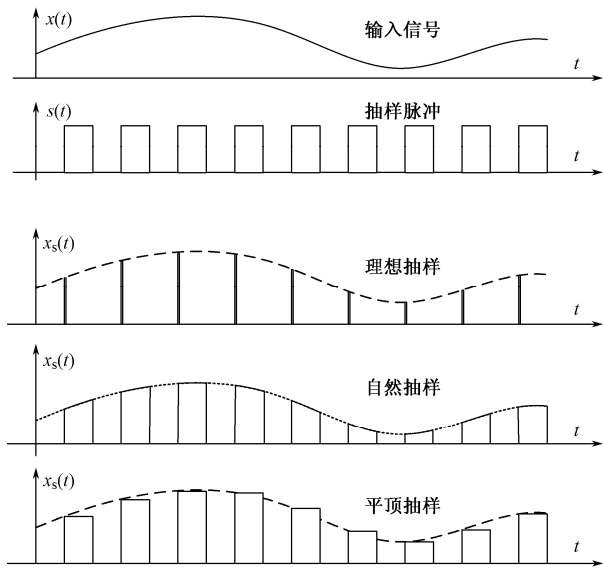


图 3-1 实际抽样的方法

4. 重要说明

- (1) 抽样是把时间连续的模拟信号变为时间上离散的脉冲信号；
- (2) 经过抽样以后的信号在时间上是不连续（离散），但它的参量（如幅度、宽度或位置）取值不是离散的，有无穷多个可能的取值，即抽样信号仍为模拟信号；
- (3) 抽样信号要变成数字信号，必须把抽样值离散化，这就是信号量化。

3.2.2 量化及量化信号与噪声功率比

1. 量化的概念与分类

量化是把一个的无限数集合（或连续值）映射到一个有限数的集合（或离散值）的过程。量化的过程犹如数学上的四舍五入，必然会产生误差，这种在量化过程中产生的误差称为量化误差，量化误差的大小与量化值的选取方法和信号的概率分布有关，**一般只能用统计平均的方法描述。**

信号量化的基本做法是，把信号幅度（抽样值）划分为若干个区间，取该区间预先规定的某个电平作为输出信号值。这里所说的区间在专业术语上称为量化层或量化级。

通常量化值的选取方法有舍去法、补足法、中间值法等。值得注意的一点是，**不同的量化值选取方法带来的误差是不一样的。**一般以中间值法产生的量化统计平均误差最小，但是在实际中，为了电路容易实现，通常在发信端采用舍去法量化，收信端再补半个量化级差，以达到量化误差最小的效果。

量化大体上可分为均匀量化和非均匀量化，绝对量化和相对量化，标量量化和矢量量化等。均匀量化是指在整个信号取值范围内，量化间隔（分层）始终相等的量化方案；非均匀

量化是指在整个信号取值范围内，量化间隔（分层）不完全相等的量化方案。

绝对量化是对抽样值绝对电平进行的量化过程；相对量化是指对相邻抽样的差值，即相对变化量，进行的量化过程。

标量量化仅对信号参量（如幅度、频率、相位等）的绝对大小量化，是一维空间信号的量化问题，语音信号一般采用此方式；矢量量化是对信号参量（如幅度、频率、相位等）的绝对大小量化和方向等的联合量化，是多维空间的量化问题，图像信号一般采用此方式。

2. 均匀量化与非均匀量化

（1）均匀量化。设信号的取值范围（0~2 048 mV），每个量化层的大小都为 1 mV，即均匀相等时，共有 2 048 个量化层或量化级，其量化值的三种选取方法举例如表 3-2 所示。

表 3-2 均匀量化时量化值的三种选取方法举例

| 量化层序号 | 取 值 范 围          | 量 化 电 平 |       |         | 量 化 间 隔 |
|-------|------------------|---------|-------|---------|---------|
|       |                  | 舍取法     | 补足法   | 中间值法    |         |
| 1     | [0~1 mV)         | 0       | 1     | 0.5     | 1       |
| 2     | [1~2 mV)         | 1       | 2     | 1.5     | 1       |
| 3     | [2~3 mV)         | 2       | 3     | 2.5     | 1       |
| 4     | [3~4 mV)         | 3       | 4     | 3.5     | 1       |
| 5     | [4~5 mV)         | 4       | 5     | 4.5     | 1       |
| 6     | [5~6 mV)         | 5       | 6     | 5.5     | 1       |
| 7     | [6~7 mV)         | 6       | 7     | 6.5     | 1       |
| 8     | [7~8 mV)         | 7       | 8     | 7.5     | 1       |
| 9     | [8~9 mV)         | 8       | 9     | 8.5     | 1       |
| 10    | [9~10 mV)        | 9       | 10    | 9.5     | 1       |
| 11    | [10~11 mV)       | 10      | 11    | 10.5    | 1       |
| 12    | [11~12 mV)       | 11      | 12    | 11.5    | 1       |
| 13    | [12~13 mV)       | 12      | 13    | 12.5    | 1       |
| 14    | [13~14 mV)       | 13      | 14    | 13.5    | 1       |
| 15    | [14~15 mV)       | 14      | 15    | 14.5    | 1       |
| 16    | [15~16 mV)       | 15      | 16    | 15.5    | 1       |
| ...   | ...              | ...     | ...   | ...     | ...     |
| 2 048 | [2 047~2 048 mV) | 2 047   | 2 048 | 2 047.5 | 1       |

（2）非均匀量化。设信号的取值范围（0~2 048 mV），最小量化层为 1 mV，最大量化层为 64 mV，非均匀量化值的 3 种选取方法举例如表 3-3 和图 3-2 所示。

表 3-3 非均匀量化时量化值的 3 种选取方法举例

| 量化层序号 | 取 值 范 围  | 量 化 电 平 |     |      | 量 化 间 隔 |
|-------|----------|---------|-----|------|---------|
|       |          | 舍取法     | 补足法 | 中间值法 |         |
| 1     | [0~1 mV) | 0       | 1   | 0.5  | 1       |
| 2     | [1~2 mV) | 1       | 2   | 1.5  | 1       |
| ...   | ...      | ...     | ... | ...  | ...     |

| 量化层序号 | 取值范围             | 量 化 电 平 |       |       | 量 化 间 隔 |
|-------|------------------|---------|-------|-------|---------|
|       |                  | 舍取法     | 补足法   | 中间值法  |         |
| 16    | [15~16 mV)       | 15      | 16    | 15.5  | 1       |
| 17    | [16~17 mV)       | 16      | 17    | 17.5  | 1       |
| ...   | ...              | ...     | ...   | ...   | ...     |
| 32    | [32~34 mV)       | 32      | 34    | 33    | 2       |
| 33    | [34~36 mV)       | 34      | 36    | 35    | 2       |
| ...   | ...              | ...     | ...   | ...   | ...     |
| 48    | [64~68 mV)       | 64      | 68    | 66    | 4       |
| 49    | [68~72 mV)       | 68      | 72    | 70    | 4       |
| ...   | ...              | ...     | ...   | ...   | ...     |
| 64    | [128~136 mV)     | 128     | 136   | 132   | 8       |
| 65    | [136~144 mV)     | 136     | 144   | 140   | 8       |
| ...   | ...              | ...     | ...   | ...   | ...     |
| 80    | [256~272 mV)     | 256     | 272   | 264   | 16      |
| ...   | ...              | ...     | ...   | ...   | ...     |
| 96    | [512~544 mV)     | 512     | 544   | 528   | 32      |
| ...   | ...              | ...     | ...   | ...   | ...     |
| 112   | [1024~1 088 mV)  | 1 024   | 1 088 | 1 056 | 64      |
| ...   | ...              | ...     | ...   | ...   | ...     |
| 128   | [1 984~2 048 mV) | 1 984   | 2 048 | 2 016 | 64      |

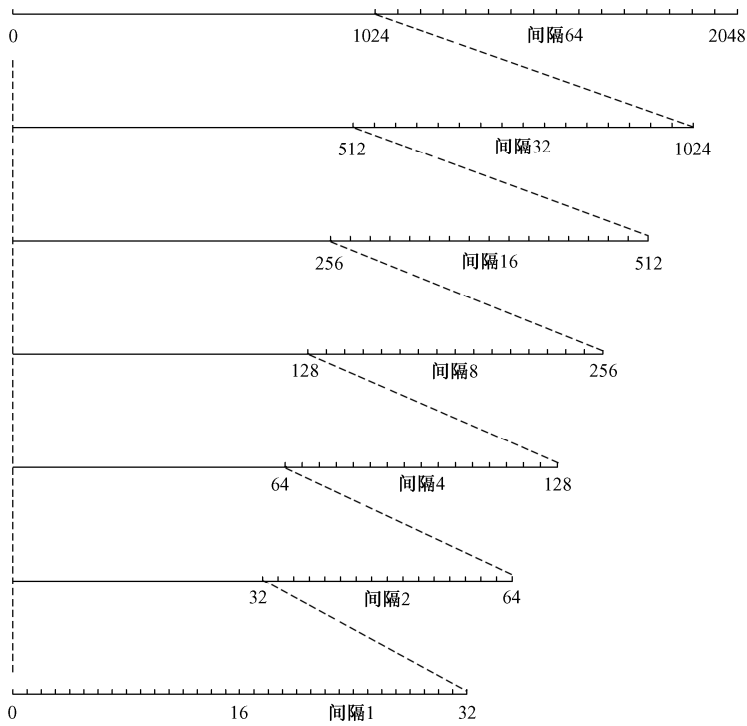


图 3-2 非均匀量化方法举例

3. 绝对量化与相对量化

如图 3-3 所示， $x(t)$ 信号  $t_1$  时刻抽样的绝对电平为 2.5 V，相邻的  $t_2$  时刻抽样绝对电平为 3.7 V，若对此绝对电平进行直接量化，就称为绝对量化过程，如分别量化为 2 和 3。

若将相邻抽样的差值， $t_2$  与  $t_1$  时刻电平差， $3.7\text{ V}-2.5\text{ V}=1.2\text{ V}$ ，即相对变化量进行量化，就称为相对量化过程，例如： $t_2$  时刻相对  $t_1$  时刻的相对变化量的量化值为 1。

在同样量化分层数情况下，相对量化的分层间隔将大大减小，这可以提高量化精度，减小量化误差。

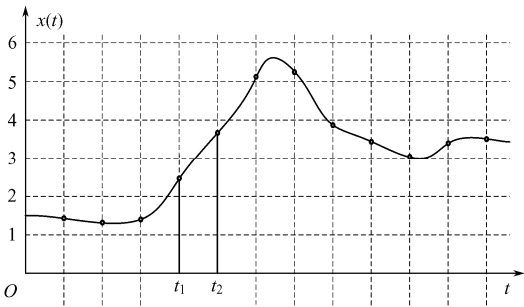


图 3-3 绝对量化与相对量化方法举例

4. 均匀量化信号与噪声功率比

设信号  $x(t)$  取值范围为  $(-a, +a)$ ，取值的概率密度分布函数为高斯分布，均值为  $a_x$ ，方差为  $\sigma_x$ ，对其均匀量化，量化间隔为  $\Delta$ ，分层（量化级）数为  $Q=2^k$ ，如图 3-4 和图 3-5 所示。

利用概率论知识通过统计数学分析，量化信号与噪声功率比为

$$S_q / N_q = 4.77 + 6k - 20\lg(a / \sigma_x) \tag{3-2}$$

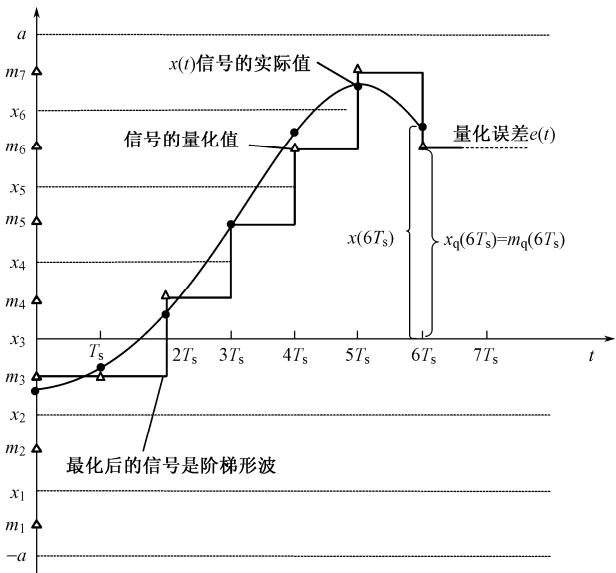


图 3-4 信号  $x(t)$  的量化过程

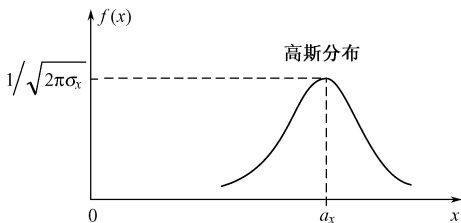


图 3-5 信号  $x(t)$  取值的概率密度分布函数

国家电话传输标准规定，满足输出信噪比要求的输入信号取样范围达到  $20\lg(a/\sigma_x) \geq 40$  dB 时，信噪比不应低于 26 dB，其量化信噪功率比也不应低于 26 dB，根据式 (3-2) 可得

$$\begin{aligned} 26 &\leq 4.77 + 6k - 40 \\ k &\geq 11 \end{aligned} \quad (3-3)$$

结论：均匀量化比特率过高，最少需要 11 位二进制编码，如果压缩编码位数或减少量化级数，则量化信噪比在输入小信号时又不能满足要求，其原因是当信号较小时， $S_q$  较小，而  $N_q$  与信号大小无关，因此小信号时  $S_q/N_q$  明显下降，这是均匀量化的致命缺点。语音信号的实际压缩编码，一般采用非均匀量化或相对量化。

### 3.2.3 编码方案及比较

#### 1. 何谓编码

编码是把一个的有限数集合（或离散值）映射到为二进制、十进制、十六进制等编码集合的过程，即把量化电平映射为代码的过程，其逆过程称为解码或译码。

#### 2. 编码的特点

(1) 在数据通信、计算机通信中，普遍采用二进制数字通信技术，所以一般信源压缩编码是把量化电平的有限数集合映射为二进制代码；

(2) 量化电平值与二进制编码不一定数值相等，但一定会一一对应；

(3) 为了学习理解方便，通常可认为量化电平序号和自然二进制编码（值）完全相等，而实际上不会相等；

(4) 量化电平数越多，需要的码位数越多；

(5) 译码是把二进制码恢复成量化电平的过程。

#### 3. 编码方案分类

理论上来说，任何一种可逆的二进制码组都可以用于编码。

常见的二进制码组有 3 种：自然二进制码组 NBC (Natural Binary Code)；格雷二进制码组或反射二进制码 RBC (Gray or Reflected Binary Code)；折叠二进制码组 FBC (Folded Binary Code)。

下面以均匀量化方案为代表，说明上述 3 种编码方案，如表 3-4 所示。



表 3-4 常见的 3 种二进制码组编码方案举例

| 量化级编号 | 电平范围  | 量化方案 1 | 量化方案 2 | 自然二进制码 | 折叠二进制码 | 格雷二进制码 |
|-------|-------|--------|--------|--------|--------|--------|
| 0     | -7~-8 | -7.5   | -8     | 0000   | 0111   | 0000   |
| 1     | -6~-7 | -6.5   | -7     | 0001   | 0110   | 0001   |
| 2     | -5~-6 | -5.5   | -6     | 0010   | 0101   | 0011   |
| 3     | -4~-5 | -4.5   | -5     | 0011   | 0100   | 0010   |
| 4     | -3~-4 | -3.5   | -4     | 0100   | 0011   | 0110   |
| 5     | -2~-3 | -2.5   | -3     | 0101   | 0010   | 0111   |
| 6     | -1~-2 | -1.5   | -2     | 0110   | 0001   | 0101   |
| 7     | 0~-1  | -0.5   | -1     | 0111   | 0000   | 0100   |
| 8     | 0~1   | 0.5    | 0      | 1000   | 1000   | 1100   |
| 9     | 1~2   | 1.5    | 1      | 1001   | 1001   | 1101   |
| 10    | 2~3   | 2.5    | 2      | 1010   | 1010   | 1111   |
| 11    | 3~4   | 3.5    | 3      | 1011   | 1011   | 1110   |
| 12    | 4~5   | 4.5    | 4      | 1100   | 1100   | 1010   |
| 13    | 5~6   | 5.5    | 5      | 1101   | 1101   | 1011   |
| 14    | 6~7   | 6.5    | 6      | 1110   | 1110   | 1001   |
| 15    | 7~8   | 7.5    | 7      | 1111   | 1111   | 1000   |

自然二进制码（NBC）就是一般的十进制正整数的二进制表示，简单直观易记，但对双极性信号编码不如折叠编码方便。

折叠二进制码（FBC）相当于计算机中符号幅度码。左边第一位表示正负号，第二位开始至最后一位表示幅度。这里第一位用“1”来表示正，用“0”表示负。其习惯恰好与计算机中的习惯相反。由于绝对值相同的折叠码，其码组除第一位外都相同，相当于相对零电平对称折叠，所以形象化地称为折叠码，其编码过程简化了一位。

格雷二进制码组或反射二进制码（RBC）简称格雷码，其特点是任何相邻电平的码组，只有一位码位发生变化，相邻两组代码的码距均为 1。

4. 编码方案比较

以上 3 种编码，自然二进制码的编码最简单，而且解码时也可逐位独立进行。格雷二进制编码较复杂，解码时不能逐位独立进行。折叠二进制码编码，除最高位码外，其余的码价是自然二进制码的折叠，因此最适合对双极性信号编码。

另外，从传输过程中误码所产生的噪声角度出发。如假定传输过程中误码是随机的，而 0—1、1—0 的误码是等概率出现的，且误码概率很小，一个码组中两位以上出现误码概率可以忽略。对上述 3 种码可以算出由误码噪声而引起的作为音量函数的信噪比，其结果如图 3-6 所示。

从图 3-6 中可以看出，由误码噪声引起的信噪比是随音量的下降而恶化的。另外，上述 3 种编码关系中，对中、高音量的信号，误码噪声的信噪比差异较小，加对低音量的信号而言其差异显著。其中折叠二进制编码要优于自然二进制编码和格雷二进制编码。

这是因为自然二进制编码和格雷二进制编码中最高位比特的误码总是对应于量化范围 1/2 处的大噪声值；而在折叠二进制编码中，最高位比特的误码相当于信号的极性反转，这

时，所产生的噪声是与音量成比例的，因而低音量时噪声也就小，这就是折叠二进制码优越的主要原因。

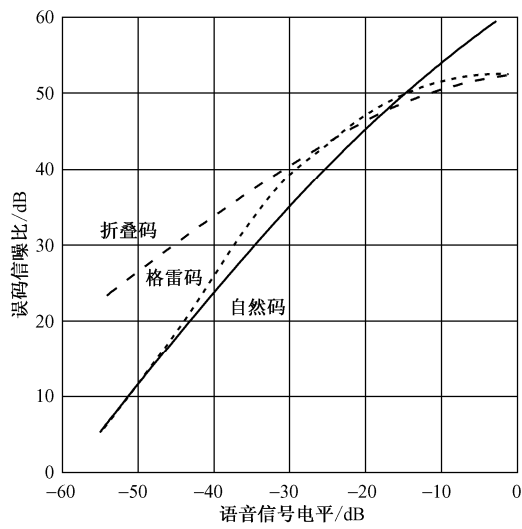


图 3-6 3 种编码方案由误码噪声而引起的作为音量函数的信噪比曲线

鉴于上述两方面的优点，目前在实际信源压缩编码方案中多采用折叠二进制码的编码方案。

3.2.4 脉冲编码调制（PCM）

脉冲编码调制（Pulse Code Modulation, PCM）简称脉码调制，是一种将模拟语音信号变换成数字信号的编码方式，如图 3-7 所示。

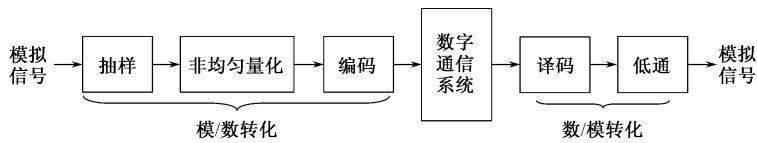


图 3-7 脉冲编码调制系统组成框图

1937 年，法国工程师 Alec Reeres 最早提出脉冲编码调制的概念。1946 年美国贝尔实验室实现了第一台 PCM 数字电话终端机。

20 世纪 70 年代后期，超大规模集成电路的 PCM 编/解码器的出现，使 PCM 在光纤通信、数字微波通信、卫星通信中获得了更广泛的应用。

PCM 指将模拟话音信号经抽样、非均匀量化、编码 3 个步骤变换为一个 64 kb/s 速率的数字信号流的过程。

1. 抽样

抽样的目的是把时间上连续的模拟信号  $x(t)$  转换成离散时间、幅度连续的抽样信号 (PAM)，为在收端能顺利恢复原信号，要求抽样脉冲频率应满足抽样定理，即抽样脉冲频率

至少要大于或等于被抽样模拟信号中所包含最高频率的两倍。对于 300~3 400 Hz 的话音信号，一般规定抽样脉冲频率为 8 000 Hz。

2. 非均匀量化

ITU-T 建议 G.712 对 PCM 设备中采用的非均匀量化与编码有明确规定，有 15 折线近似  $\mu$  律的和 13 折线近似 A 律，我国采取的是后一种。

13 折线近似 A 律的非均匀量化方案，首先将信号取值范围（-2 048~+2 048）按 1/2 倍（或 2 倍）划分为 16 段，然后每段又划分为 16 个量化级（分层），这样，整个信号范围有 16 段×16 量化级=256 个量化级。

信号取值范围在（0~+2 048）时，8 段×16 量化级=128 个量化级情况，如图 3-8 所示。

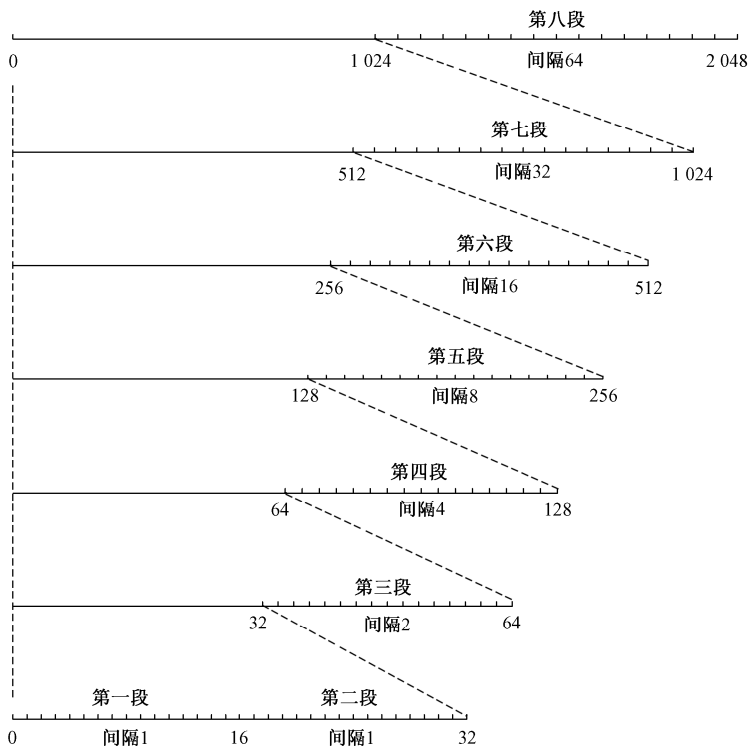


图 3-8 13 折线近似 A 律的非均匀量化方案

3. PCM 编码

ITU-T 建议 G.712 对 13 折线近似 A 律编码的规则是这样的：用 8 位码组表示一个量化幅值，可共代表 256 个幅度化量级。对 8 位码组的分配是：

$$\underbrace{M_1}_{\text{极性码}} \quad \underbrace{M_2 M_3 M_4}_{\text{段落码}} \quad \underbrace{M_5 M_6 M_7 M_8}_{\text{电平码}}$$

第 1 码位表示幅值的符号，正电平为“1”，负电平为“0”；第 2~8 码位表征幅值的绝对值大小，第 2~4 码位表示所处段，第 5~8 码位表示所在段的幅值。

4. 译码与低通

当编码信号通过数字通信系统，传输到接收端，接收端要完成的主要工作是译码与低通，其中译码是编码的逆过程，将接收到的二进制码组转化为对应时刻的量化电平，这些量化电平脉冲序列再通过低通滤波器平滑后，就是近似发端信号。

3.2.5 自适应差值脉冲编码调制（ADPCM）

自适应差值脉冲编码调制（Adaptive Differential Pulse Code Modulation，ADPCM）是一种为提高数字通信系统的频带利用率而发展起来的语音压缩编码技术，ITU-T 为此制定了 G.721 推荐标准，它可在 32 kb/s 数码率上达到 64 kb/s 的 PCM 数字电话语音质量。

ADPCM 是在差分脉码调制（DPCM）的基础上发展起来的，所以首先介绍 DPCM 的原理。

1. 差分脉码调制（DPCM）

语音信号和图像信号经抽样后得到样值序列，经分析可知，当前时刻的样值与前面相邻的若干时刻的样值之间有明显的关联。这样，可以根据前一时刻的样值来预测当前时刻的样值，预测值和实际值之差为差值。

大量统计的结果是，在大多数时间内，信号本身的功率比差值的功率要大得多，如果只传送这些差值来代替信号，那么码组所需的位数就可以显著减小。差分脉冲调制就是利用样值之间的关联进行高效率波形压缩编码的一种典型方法。

图 3-9 是差分脉码调制（DPCM）原理组成框图，图中模拟信号  $x(t)$  和预测信号  $x_q(t)$  输入到比较器，输出误差  $e(t)=x(t)-x_q(t)$ 。对  $e(t)$  进行抽样、量化和编码得到了 DPCM 信号，这是一个二进制的数字序列。此信号分成两路，一路送至信道上传输，另一路送至反馈支路，经过本地译码器译码产生预测信号。在接收端，可以用相同的译码器译码，并经过低通滤波器后恢复模拟信号。

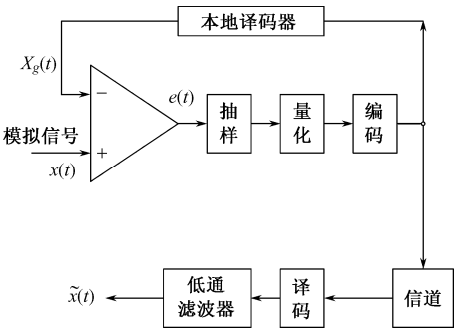


图 3-9 差分脉码调制（DPCM）原理组成框图

DPCM 与 PCM 的区别是：在 PCM 中是对信号的抽样值直接进行量化、编码后传输，而 DPCM 则是对信号抽样值与信号预测值的差值进行量化、编码后传输。由于差值信号的动态范围一般比信号小，如果输入信号统计特性已知，则进行适当预测可使差值信号范围更缩小。

实验表明，在较好图像质量的情况下，每一抽样值只需 4 bit 就够了。因此大大压缩了传

送的比特率。另一方面，如果比特速率相同，则 DPCM 比 PCM 信噪比可改善 14~17 dB。

DPCM 的缺点是较易受到传输线路噪声的干扰。因为 DPCM 能压缩比特率的实质，是由于声音和图像信号相邻的抽样值之间，存在明显的相关性，因此用一般的 PCM 传输时，信号含有多余信息。DPCM 预测减少了多余信息，所以抗传输噪声的能力必然降低。

DPCM 预测效果与信号统计特性有密切关系，要使声音和图像信号（统计特性随时间变化）获得最佳的效果，预测电路应跟踪信号性质的变化。

## 2. 自适应差分脉码调制（ADPCM）

为了能进一步提高 DPCM 方式的质量，在语音相当宽的变化范围内仍能得到最佳的性能，特别是适应因人、因时不同的语音信号变化，在 DPCM 中增加自适应系统，这里的自适应应包括自适应预测和自适应量化，增加自适应系统 DPCM 的称为自适应差分脉冲编码调制（ADPCM）。

### 1) 自适应量化

自适应量化是对 DPCM 误差  $e(t)$  信号非均匀量化过程中，让量化间隔大小的变化与输入信号方差相匹配，即量化间隔随输入信号的方差而变化。

现有的自适应量化方案有两类：一是其输入信号幅度或方差，由输入信号本身估算，这种方案称为前馈自适应量化器；二是其量化阶距，根据量化器的输出来进行自适应调整，或等效地用输出编码信号进行自适应调整，这类自适应量化方案称为反馈自适应量化器。

无论采用哪种自适应量化方案都可以改善动态范围及信噪比，比相同电平数固定量化改善 10~12 dB。

反馈控制的优点是：量化阶距的信息由码字序列提取，不需要传输或存储额外的阶距信息，但在重建输出信号时，传输误码对质量的影响比较明显。

前馈控制时，要求码字与阶距一起传输，以用来得出信号，虽然增加了复杂性，但它有可能在差错控制保护下传输阶距，减少传输误码的影响，即使在高误码率传输时，也有较好的输出信号质量。

### 2) 自适应预测

自适应预测是对 DPCM 预测信号  $x_q(t)$  进行自适应预测，并将根据前面一个抽样，确定下个预测值称为一阶预测，采用前面若干个样值，作为参考来推算预测值称为高阶预测。

为了在接收端，根据差值的编码产生下一个输入样值的准确预测（估计），可以对前面所有样值的有效信息冗余度进行加权求和，加权系数又称为预测系数。

自适应预测是使预测系数的改变与输入信号幅度值相匹配，从而使预测误差为最小值，这样预测的编码范围可减小，可在相同编码倍数情况下提高信噪比。

### 3) 32 kb/s ADPCM 系统

ITU-T G.721 建议提出了与现有 G.712 建议 64 kb/s PCM 系统兼容的 32 kb/s ADPCM 的实现，其主要技术指标满足 G.712 对语音质量要求。

32 kb/s ADPCM 编码器的方框图如图 3-10 所示。它由 PCM 码/均匀量化线性码间的变换

器、自适应量化器、自适应逆量化器、自适应预测器和量化尺度适配器组成。

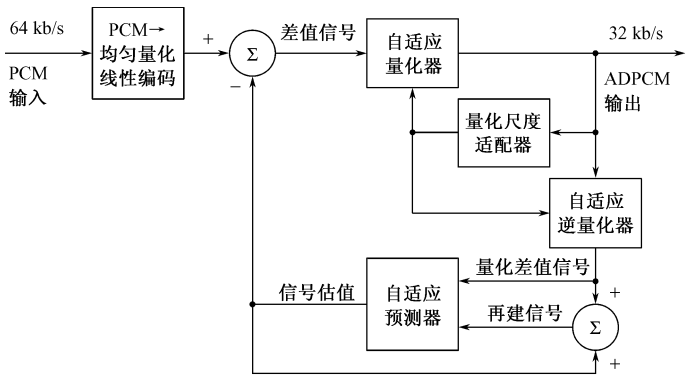


图 3-10 ADPCM 编码器简化方框图

编码器输入的信号为非均匀量化 PCM 码，可以是 A 律和μ律 PCM 码。为了便于进行数字信号运算处理，首先将 8 位非均匀量化码变换为 12 位均匀量化码，然后进入 ADPCM 部分。

均匀量化的 PCM 信号与预测信号相减获得预测误差信号。自适应量化器将该差值信号进行量化并编成 4 位 ADPCM 码输出。

ADPCM 码流通过自适应逆量化器产生量化差值信号。量化差值信号与信号预测值相加形成重建信号。自适应预测器对重建信号及量化差值信号进行运算形成对输入 PCM 信号的预测信号估值。

量化尺度适配器包括定标因子自适应和自适应速度控制两电路。编码器中的量化器和逆量化器的自适应均受量化尺度适配器中的定标因子的控制。

32 kb/s ADPCM 译码器的方框图如图 3-11 所示。它由自适应逆量化器、自适应预测器、均匀量化线性码/PCM 码变换器、量化尺度适配器以及同步编码调整器组成。译码器中有与编码相同的电路，只是多了一个同步编码调整器，其作用是使级联工作时不产生误差积累。

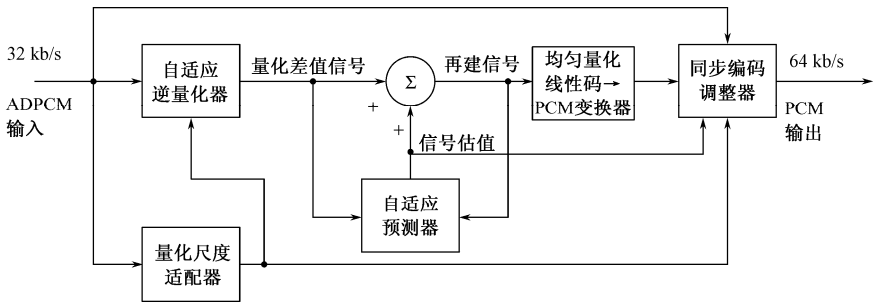


图 3-11 32 kb/s ADPCM 译码器方框图

3.2.6 子带编码（SBC）

把语音信号的频带分割成不同的频带分量（称为子带），然后再分别对这些子带独立地进行抽样编码的方式，称为子带编码（Sub-Band Coding，SBC）。

子带编码将信号分解成不同频带分量的过程去除了信号的冗余度，得到了一组互不相关的信号。这同 DPCM 方式的机理虽然不同，但从去除冗余度角度这两者又是相似的。

SBC 主要有两个优点：

首先是可以通过分配给各子带不同的量化间隔和编码比特数来分别控制它们的信噪比，以较低的总码率获得较好的语音质量。这主要靠利用人耳听觉的“掩蔽”效应实现的。例如，语音信号低频部分涉及语音基音周期和第一共振峰等，对语音清晰度等主观品质影响较大，信噪比应高些，即量化间隔选小些，分配较多的编码比特数；对语音的高频部分，量化噪声对语音质量的影响小些，信噪比可低些，所以量化间隔可以稍大些，用较少的比特编码。这样，便可在保证语音质量的前提下，使编码的总比特数降低。基于此原因，子带编码方式既不是纯粹的波形编码又不是纯粹的参量编码，它是二者的结合，应属于混合编码，这类编码方式也称为频域编码。

其次，子带编码各子带的量化噪声相互独立，被束缚在自己的子带内，这样就能避免输入电平较低的子带信号被其他子带的量化噪声所淹没。

实现子带编码原理的系统组成方框示意图如图 3-12 所示。

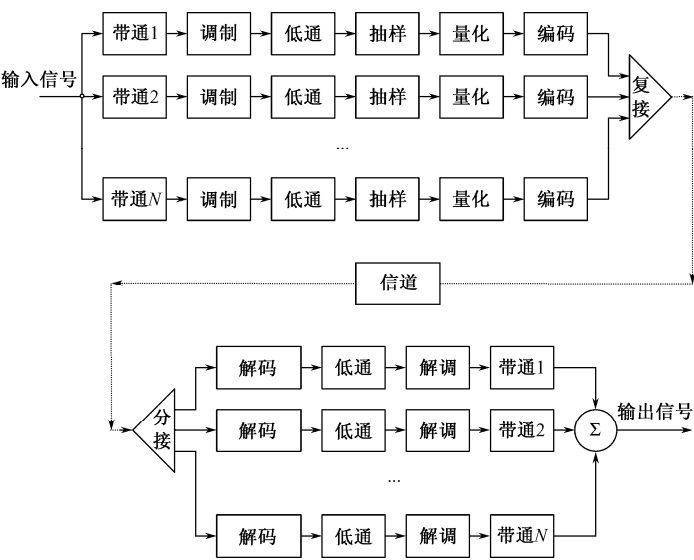


图 3-12 子带编码系统组成方框示意图

在发端，首先通过带通、调制将各个子带信号转换成低通信号，然后再根据低通抽样定理对各子带信号进行抽样，由于子带信号转换成了低通信号，所以可使抽样频率降低到各子带带宽的两倍。接着分别量化编码，最后将各路数字流，合在一起，送往信道中传输。

在收端，再将它分接（即分路）、解码并恢复各个子带信号，然后将各个子带信号解调，组合起来还原成原语音信号。

实验证明，16 kb/s SBC 系统的语音质量相当于 24 kb/s 的 ADPCM 系统，其典型参数如表 3-5 所示。

表 3-5 16 kb/s SBC 系统的典型参数

| 子带序号    | 1           | 2         | 3       | 4       | 5       |
|---------|-------------|-----------|---------|---------|---------|
| 频率范围/Hz | 1 600~3 200 | 800~1 600 | 400~800 | 200~400 | 100~200 |
| 编码比特/样值 | 2           | 2         | 4       | 5       | 5       |

子带编码器可应用于模拟线路，并兼容传输数字化语音，和在一般线路上传输数字语音。目前有人还提出了可变子带编码器，其比特速率最低可达 4.8 kb/s，其语音质量可与 7.2 kb/s 的固定子带编码相比拟。

### 3.3 语音参量压缩编码技术

参量压缩编码的原理和设计思想与波形压缩编码完全不同。

波形压缩编码的基本思路是忠实地再现语音的时域波形，为了降低比特率，可充分利用抽样点之间的信息冗余性对差值信号进行编码，在不影响语音质量的前提下，比特率可以降低至 32 kb/s。

参量压缩编码根据对语音的形成机理分析，着眼于构造语音生成模型，要了解参量压缩编码原理，首先必须了解语音形成机理，以及语音产生的数学模型。

#### 3.3.1 语音产生的过程及数学模型

语音是由人类的发声器官产生，气流通过声带、声门，由咽腔、口腔和鼻腔等共同控制而发音，完整的发声系统还包括由肺、支气管、气管组成的次声门系统，次声门系统是产生语音能量的源泉。当空气从肺中呼出来时，气流由于声道某一地方的收缩而受到扰动发出声波。语音形成的大致过程可如图 3-13 所示。

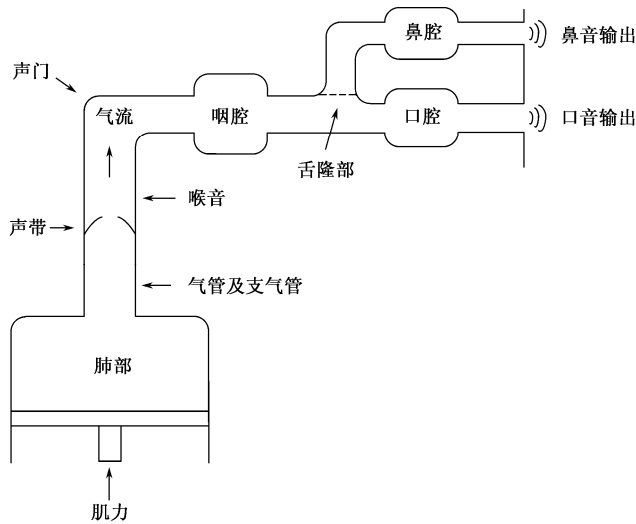


图 3-13 语音形成的大致过程

语音按其激励分为 3 种形式：浊音、清音和爆破音。

#### 1. 浊音与基音

浊音又称有声音，伴有声带振动的音，发浊音时声带在气流的作用下准周期地闭合或开启。即气流通过声门、声带的张力使得声带发生张弛式的振荡，这时就会产生一股准周期气流，这股气流激励声道时就产生浊音，如图 3-14 所示。



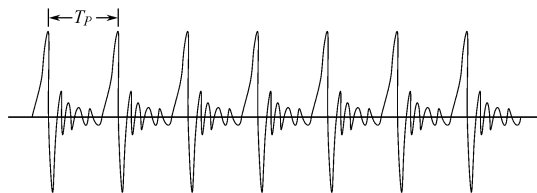


图 3-14 浊音

由于声带振动有不同的频率，因此浊音就有不同的音调，称之为基音频率，这个倒数称为基音周期，若基音周期为  $T_p$ ，则基音频率为  $f_p$ 。基音频率与个人声带有关，在很大程度上反映了个人的特征。一般而言，男性基音频率大致分布在  $50 \sim 250 \text{ Hz}$  左右，女性和小孩的基音频率在  $100 \sim 400 \text{ Hz}$  左右。

## 2. 清音和爆破音

清音又称无声音，声带不振动的音，由声学 and 流体力学知，当气流速度达到某一临界速度时，就会引起湍流，此时声带不振动，声道被噪声状随机波激励，产生较小幅度声波，其波形与噪声很像，这就是清音，清音信号没有准周期特性，如图 3-15 所示。

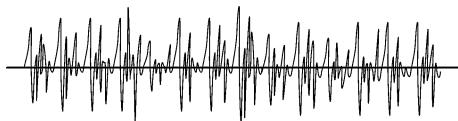


图 3-15 清音

如果声道完全闭合，在闭合后产生气压然后突然释放，这就得到了爆破音。

从语音信号的频谱分析可知，语音信号除基音外还存在基音的多次谐波，浊音信号的能量主要集中在各基音谐波的频率附近，而且主要集中于低于  $3 \text{ kHz}$  的范围。

另外，气流压的不同强度就对应为声音的音量大小。

## 3. 共振峰及声道模型参数

人的声道和鼻道都可以看成非均匀截面的声道管，如图 3-16 所示，声道管的谐振频率称为共振峰频率。共振峰频率和声道的形状和大小密切相关，每种形状都有一套共振峰频率作为其特征，改变声道的形状参量（数） $a_1, a_2, a_3, \dots, a_N$ ，就产生不同的声音。

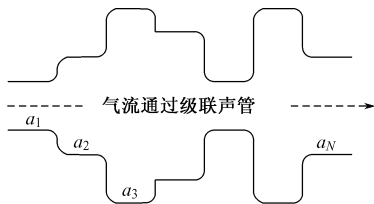


图 3-16 声道模型（口腔和鼻腔）

由流体力学知，声道频率特性（唇口声速与声门声速之比）与谐波类似，频率特性对应的谐振点叫做共振峰频率，共振峰只会出现在浊音频谱中。

共振峰频率由高到低依次为第一共振峰，第二共振峰，第三共振峰…，分别用  $F_1$ ， $F_2$ ， $F_3$ …表示。一般浊音有 5 个共振峰，前 3 个最重要。

4. 语音信号产生的数学模型

根据以上对实际发音器官和发音过程的分析，可将语音信号发生过程抽象为如图 3-17 所示的数学模型，也就是说，语音是语音激励源激励一个参数变化的声道而产生的。这里，以具有一定周期的脉冲源表示浊音的激励，以分布平坦的噪声源来表示清音激励。

周期信号源表示浊音激励源，随机噪声信号源表示清音激励。根据语音信号种类，由浊/清音开关决定连接哪一种激励源。

声道模型（口腔和鼻腔）可以看成是一个时变线性系统对声波系统起滤波作用，其参数随口腔运动缓慢变化，增益控制代表语音的强度。

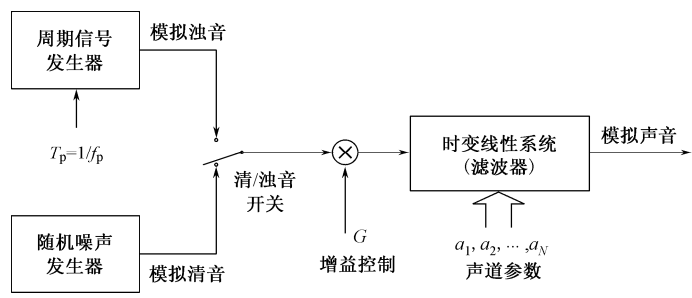


图 3-17 语音信号产生的数学模型

3.3.2 参量编码与声码器简介

语音信号的特征参数主要有：基音周期、共振峰频率、清/浊音判决和语音强度。基音周期和清/浊音判决可以同时获得，其方法主要有三大类：

- （1）时域法，指直接用语音信号波形来估计的方法；
- （2）频域法，指将语音信号变换到频域来估计的方法；
- （3）混合法，指综合利用语音信号的频域和时域特性来估计的方法。

语音信号参数编码始于 1939 年美国人 Homer Dudgey 发明的声码器，它是根据语音信号的特征参数来编码，又称声源编码，也称为声码器编码，又叫“声码器技术”。

参数编码是以发音机制的模型为基础，用一套模拟声带频谱特性的滤波器系数和若干声源参数来描述这个模型，在发送端从模拟语音信号中提取各个特征参量并进行量化编码，从而实现语音信号到数字信号的转变的。

构成声码器的主体是一个滤波器，这个滤波器的作用相当于人类的发音器官——喉、嘴、舌的组合。声码器中滤波器的系数和若干声源参数，由语音信号的频谱特性所决定。

在发送端，对输入语音信号加窗（截取一帧语音），对该帧语音信号进行分析，模拟发声结构，求出这帧语音信号的激励参数（包括清浊音判决信息和基音周期参数）和声道参数。然后，对语音信号的特征参数进行提取和编码，最后输出。

在提取语音信号的特征参数时，利用语音信号生成模型，幅度谱上要逼近原语音，并在编码前进行清浊音判断，力图使重建语音信号具有尽可能高的可懂度，即保持原语音的语意。

在接收端，根据接收到的参量编码和上述模型还原生成发话者的音素。即在浊音段，激励信号采用具有该浊音段基音周期的周期脉冲序列；在清音段，激励信号采用白噪声序列。在合成端将激励信号通过代表声道特性的时变线性系统滤波器产生合成语音。

具体来说，用上述模型模拟语音的发声过程，语音信号可以看做不同激励信号通过时变线性系统产生的输出。发送端对该模型的参量（参数）进行编码，相当于对该语音信号的主要特征而并非具体的语音波形幅值编码。

根据分析可知，模型参数的更新频度较低，并可利用抽样值间的一定相关性，故可有效降低编码比特率。

声码器主要有通道声码器、共振峰声码器、相位声码器、线性预测（LPC: Linear Predictive Coding）声码器等，LPC 声码器是重要的一种。

其中 LPC 声码器是以线性组合模型均方误差最小意义下逼近原始波形的方法提取参数，较好地解决了编码速率和语音质量的问题，以其成熟的算法和参数的精确估计成为研究的主流，并已走向实用。美国政府 1980 年公布的 2.4 kb/s 线性预测编码算法 LPC-10 就是采用的这种方法。1986 年，美国第三代保密电话装置采用了 2.4 kb/s 的 LPC-10e（LPC-10 的增强型）作为语音处理方法。

### 3.3.3 线性预测声码器的原理

下面首先介绍预测编码的概念，然后再说明线性预测声码器的原理。

#### 1. 预测声码器（预测编码）的基本概念

在前面讨论 ADPCM 时曾指出，利用信号波形的抽样值之间的相关性，采用传输信号差值编码方式，来降低编码速率，实际上 ADPCM 就是一种预测编码方式。

预测编码的基本思路不是直接对信源输出信号进行编码，而是充分利用信源信号各分量之间的统计相关性，先将信源输出信号进行预测变换后，再对信源输出与预测值的差值进行编码，其原理图如图 3-18 所示。

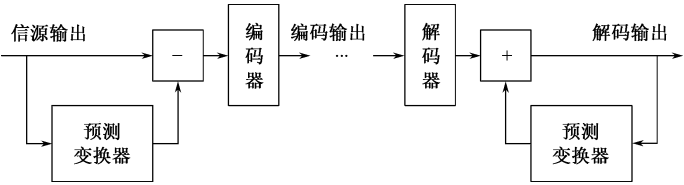


图 3-18 预测编码的基本概念

在参量编码中，可利用语音信号的特点，发送端仅对发音机制的模型预测参数和声音特征参量进行编码传输，这就是预测声码器的基本思想。

若将预测误差当做白噪声（清音）或周期脉冲（浊音），由接收端自行产生合成而无须传送，预测变换器是按线性特性进行预测变换，即采用线性预测编码（Linear Predictive Coding, LPC），则称为 LPC 声码器，如图 3-19 所示，注意其为开环结构。

若将预测误差（或残差）信号经量化编码后发送给对方，直接利用残差信号去激励。这种 LPC 编码器称为残差激励线性预测编码器（Residual-Excited Linear Prediction Coding，

RELPC)，如图 3-20 所示。

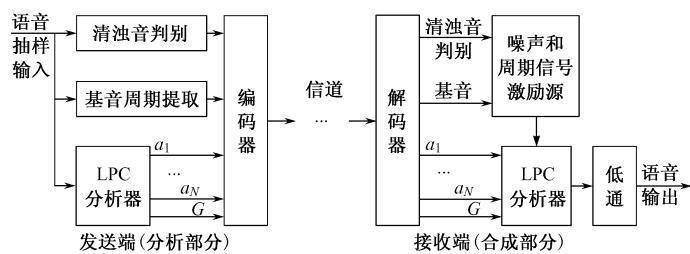


图 3-19 LPC 声码器方框示意图

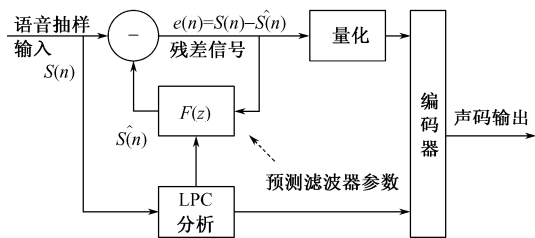


图 3-20 残差激励线性预测编码器

在这种编码器中，可通过 LPC 的分析将滤波器参数求出，并利用这些参数决定的预测滤波器  $F(z)$  得出  $\hat{S}(n)$ ，再与信号  $S(n)$  相减，就得到残差信号  $e(n)$ 。将  $e(n)$  量化后与滤波器参数编码一起送到对方，就可通过译码而得到这些参数，然后再利用这些参数得到合成的语音。

线性预测的目的是使原始信号与估计信号的误差（称为残差信号或者余量信号）接近于白噪声。残差信号的功率越小，预测滤波器越好。一般提高滤波器的阶数可以减小残差信号的功率，但对于特定的语音帧而言，当阶数达到一定值，继续提高滤波器的阶数，残差信号的功率并没有明显的减小。

## 2. 语音 LPC 的实现

通过前面分析，我们已知道，语音中的浊音有一定的周期性，识别出这种周期性用来编码，可大大压缩码率。

浊音的基音频率一般为 100~400 Hz，按标准的 8 kHz 取样率，在一个基音周期内有 20~80 个样点值。

因为周期函数的自相关函数  $R(\tau)$  也是周期的，且在  $\tau=T$ （周期）时达到最大，因此，如果我们计算浊音的  $R(\tau)$ ，则在  $\tau=20\sim80$  个样点之间  $R(\tau)$  将会出现峰值，两峰值的间隔就是基音周期。但是并非所有浊音都能产生如图 3-14 所示的稳定易识的基音周期。如果检测不准，可能产生奇怪的“非人音”，所以我们每 10~20 ms 就寻找一次基音周期。

至于清音，根据统计， $\tau>12$  个样点后  $R(\tau)$  就很小了，这意味着清音的关联长度主要在 12 个样点的抽样值以内，一般保留  $N=8\sim10$  个预测系数就够了。

综上所述，LPC 声码器的主要实现步骤如下：

- （1）首先对语音信号加窗，即把语音每 80 或 160 个抽样值（10~20 ms）分成一帧；
- （2）对本帧语音经过清浊音判决提取清浊音判决信息；

- (3) 在  $\tau=20\sim80$  个样点范围内计算  $R(\tau)$ ，利用  $R(\tau)$  两个峰值的间隔找出基音周期  $\tau$ ；
- (4) 清音固定使用  $N$  个预测系数  $a_1, a_2, \dots, a_N, N=8\sim10$ ；
- (5) 按最小均方误差准则计算最佳线性预测系数  $a_1, a_2, \dots, a_N$ ；
- (6) 将全部预测参数  $a_1, a_2, \dots, a_N$ ，连同清浊音判决信息、基音周期和预测误差  $e_k$ ，经过量化编码后发送出去；
- (7) 在接收端，通过参数译码得到这一帧语音信号的特征参数，包括基音周期、清浊音判决信息、增益控制参数和预测滤波器系数，将这一组参数作用于语音合成滤波器，再经过 D/A 转换器就得到合成语音信号。

由此可见，语音的 LPC 同样可用基本 DPCM 系统实现。差别仅在于预测器参数每帧更新一次并传至接收端，而且为了保证在计算一帧参数时仍能连续采集下一帧信号，需要设置两块存储器来分别缓存前、后两帧的语音样本，且参数必须在一帧时间内算出。由于既要传送多个参数，又要传送预测误差，所以码率不能太低，一般用于 9.6~19.2 kb/s 的中速语音编码。

为了进一步降低码率，我们可以不传误差只传参数，由接收端，利用这些参数自己合成语音，如图 3-21 所示。

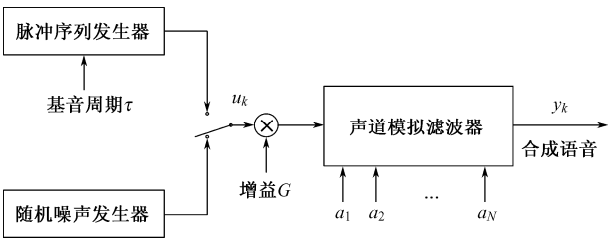


图 3-21 LPC 声码器语音合成框图

这种方法就是前述的 LPC 声码器，其原理很简单：若是浊音，采用重复周期为  $\tau$  的脉冲串作为激励源；若是清音，则采用白噪声作为激励源，至于声道模拟器，其实就是 DPCM 系统的接收部分。

不过声码器的两种激励源，其平均幅度都是固定的，不能反映话音大小的相对变化。为了克服这个缺点，还需要用一个增益  $G$  来控制激励源输出的大小。如果用  $u_k (k=0, 1, \dots, N)$  表示激励源的输出幅度，那么直接令  $u_k$  的平均能量等于预测误差  $e_k$  的平均能量，可以认为是合理的。因此  $G$  可由下式得到：

$$G \sum_{k=0}^N u_k = \sum_{k=0}^N e_k \tag{3-4}$$

而 LPC 声码器输出的合成语音可表示为

$$y_k = \sum_{i=1}^N a_i \cdot y_k + G \cdot u_k \tag{3-5}$$

LPC 声码器计算出来的各种参数  $G, \tau, a_1, a_2, \dots, a_N$  都是模拟量（ $\tau$  除外），还必须先量化再编码。各种参数的范围及影响不尽相同，实用中还希望总码率尽量靠近  $150 \times 2^n (0, 1, 2, \dots) \text{ b/s}$  的典型数传码率。

线性预测语音合成器利用线性预测语音分析方法，通过分析自然语音样本，计算出 LPC

系数, 就可以建立信号产生模型, 从而合成出语音。线性预测合成模型是一种“源—滤波器”模型, 由白噪声序列和周期脉冲序列构成的激励信号, 经过选通、放大并通过时变数字滤波(由语音参数控制的声道模型), 就可以再获得原来的语音信号。

**例 5-1** 一个 10 阶 LPC 声码器, 每帧用 48 bit 量化, 若按如下方案分配给各个参数:

|      |   |        |       |       |       |       |       |       |       |       |       |          |
|------|---|--------|-------|-------|-------|-------|-------|-------|-------|-------|-------|----------|
| 参 数  | G | $\tau$ | $a_1$ | $a_2$ | $a_3$ | $a_4$ | $a_5$ | $a_6$ | $a_7$ | $a_8$ | $a_9$ | $a_{10}$ |
| 量化位数 | 5 | 5      | 5     | 4     | 4     | 4     | 4     | 4     | 4     | 4     | 3     | 2        |

则总的数码率为 48 bit/帧 $\times$ 50 帧/s=2 400 b/s=150 $\times$ 24 b/s。

一种开环的 LPC 编码, 方法简单, 合成语音波形与原语音波形相差较大, 故语音质量不高。

### 3.4 混合压缩编码技术

混合编码是介于波形编码和参量编码之间的一种编码方法, 它综合了波形编码和参量编码的长处, 兼有参量编码低速率和波形编码高质量的优点。

实现混合编码的基本思想是以参量编码原理, 特别是以 LPC 原理为基础, 保留参量编码低速率的优点, 并适当地吸收波形编码中能反映波形个性特征的因素, 重点改善语音的自然度。

决定混合编码性能的主要参量有 4 个: 数据比特率、语音质量、复杂度与时延。利用技术的不断进步, 可以通过增大硬件复杂度, 改进算法并减小软件实现的复杂度, 在此基础上换取一定的语音质量与时延指标要求下的数据比特率的降低。

改进 LPC 主要从 3 个方面入手:

- 改进语音生成物理模型、激励源结构和合成滤波器结构, 提高语音质量;
- 改进参量量化和传输方法, 进一步压缩传输速率;
- 采用自适应技术, 进一步解决系统与信源和信道之间的统计匹配。

改进 LPC 性能的第一个措施是采用新的编码方案:

(1) 余数激励声码器 (RELPC) 方案。该方案将语音基带低频信号中的一小部分 (称为余数信号) 按取样率量化后送至接收端, 以改善语音自然度。

(2) 多脉冲激励 LPC 方案。人们通过对语音的深入分析发现, 采用清浊二元模型过于简单, 因为实际上有相当部分的语音是介于两者之间的混合型。另外, 浊音激励仅采用单个脉冲源也过于简单化。为了改进上述特性, 人们提出了多脉冲激励线性预测编码 (MPLPC) 方案和规则脉冲激励长期预测编码 (RPE—LTP) 方案, 后者就是 GSM 中采用的编码方案。

(3) 码激励线性预测编码 (CELP) 方案。CELP 中应用了矢量量化技术, 它是 IS-95 与第三代移动通信系统 CDMA2000 及 WCDMA 语音编码技术的基础。

改进 LPC 性能的第二个措施, 是进一步减小和压缩参量量化后的信息速率, 主要采用矢量量化、变换和优化等技术。

改进 LPC 性能的第三个措施是采用自适应技术。自适应技术包括信源特性参量的自适应以及与信道特性相匹配的传输速率的自适应。

下面简单介绍多脉冲激励线性预测编码 (MPLPC)、规则脉冲激励长期预测编码 (RPE-LTP) 和码激励线性预测编码 (CELP) 方案。

3.4.1 多脉冲激励线性预测声码器（MPLPC）

1982 年，Atal 和 Remde 提出了多脉冲激励线性预测编码 MPLPC（Multi-Pulse Linear Predictive Coding）方案。这种算法在一帧中提取  $K$  个典型脉冲作为激励信号，由这种算法合成的语音有较好的自然度，而且有一定的抗噪能力，但是由于其采用一种准最优顺序优化的方法得到激励脉冲的位置和幅度，运算量相当大。

线性预测残差中的小信号对合成语音的质量影响不大，如果对残差信号中幅度低于某一门阈值的所有信号设为零，则适当调整阈值可以使残差信号中大多数采样点为零，用余下的幅度较大的信号作为激励信号源，其合成语音并未产生明显的畸变。在多脉冲线性预测编码中，规定在一定的时间间隔中只能出现少数个非零脉冲，然后优化每个脉冲的幅度和位置，最后以优化后的脉冲序列表示残差信号，并作为滤波器的激励源。

在 MPLPC 中，无论是清音还是浊音，都是用一帧中的有限个脉冲经过最优估值后作为激励信号源。图 3-22 给出这个编码方案的原理框图。其中，原始语音信号  $S(n)$  以帧为单位进行处理，帧长通常为 10~20 ms。对每帧原始语音首先用 LP 分析方法算出 LP 系数  $\{a_i\}$ ，然后在当前帧范围内每 5 ms 或 10 ms 用合成分析法估计一次激励脉冲，图中虚线部分是合成器。将激励脉冲输入合成器，得到合成语音  $\hat{s}(n)$ 。将合成语音与原始语音相减，得到误差信号  $e_s(n)$ ，并输入到感觉加权滤波器  $W(z)$ ，得到输出加权误差信号  $e(n)$ 。根据最小加权均方误差准则，分析出一组位置、幅度最佳的激励脉冲，然后与 LP 参数一起编码送入信道。

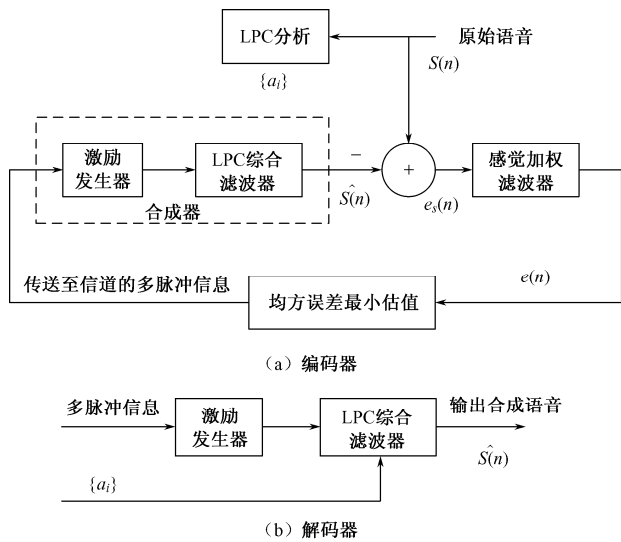


图 3-22 多脉冲激励线性预测声码器原理框图

MPLPC 的关键问题是，如果一帧中只允许用  $M$  个激励脉冲，如何求出这  $M$  个脉冲在帧内的位置与幅度，使得合成语音与原语音的感觉加权均方误差最小。

3.4.2 规则脉冲激励线性预测声码器（PRE-LPC）

PRE-LPC 由 Ed. F. Deprettere 和 Peter Kroon 在 1985 年 IEEE ICASSP 年会上首先提出。

1986 年 K. Hellwig, R. Hojmann 和 P. Wary, R. J. Sluyter 等人在 P. Kroon 提出的规则脉冲激励编码的基础上, 将它简化成实时算法, 并用 3 片 DSP 片实现了编、译器, 编码速率是 16 kb/s。以后他们与 C. Galand, M. Rosso 等人合作, 改进算法, 加入了长时预测(Long Term Prediction, LPT), 并使编码速率降低为 13 kb/s, 形成长时预测规则脉冲激励 (RPE-LTP) 编码方案。它的特点是算法简单, 语音质量达到了通信等级。

RPE-LTP 在 1988 年被确定为泛欧标准全速语音编码方案, 称为 GSM 标准。

RPE-LPC 原理与 MPLPC 有相似之处, 它们都是用搜索出来的激励脉冲序列作为激励信号源。图 3-23 是它的编、解码器示意图。

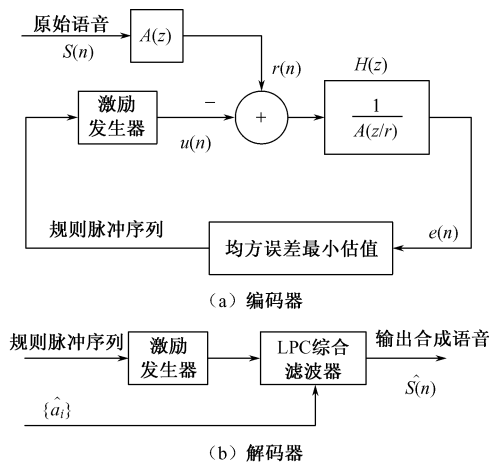


图 3-23 规则脉冲激励编解码器原理框图

其中, 语音信号经过  $P$  阶逆滤波器  $A(z)$  后, 得到余量信号  $r(n)$ 。 $u(n)$  表示激励序列,  $r(n)$  与  $u(n)$  的误差输入到综合加权滤波器  $H(z)$ , 输出综合加权误差  $e(n)$ 。优化的过程就是调整激励序列  $u(n)$ , 使  $e(n)$  在一定范围内所取的平方和最小。

### 3.4.3 码激励线性预测声码器 (CELP)

码激励线性预测编码 CELP 是利用码本 (Code Book) 来做激励源的编码方法, 即把残差信号可能出现的各种经量化了的样值, 按照一定的规则排列在存储器中, 每一样值序列有一个地址码, 这个存储器称为码本。收发双发各有一个同样的码本。

在线性预测时, 对于残差信号, 并不传输它本身, 而是先在本方的码本中检查出与这个信号最接近的样值组合的地址码, 然后将这个码本地址码经传输电路发送到对方。因为双方有同样的码本, 接收方收到这个地址码后, 就可从码本中取出这个地址的残差信号来加到滤波器上以重建语音。

因为所传的不是残差信号本身, 而只是码本上的地址码, 所以可大大减小要传送的比特数, 故可得到低速率的编码器。只要码本编得好, 既有足够的数量, 又和实际的信号极为接近, 则在较低的码速率下, 也可以得到较好的语音质量。所以这种编码方法关键是如何编好一个码本。

对码本的要求如下:



- (1) 码本中的信号应与实际信号相近，或相差很小；
- (2) 在满足条件（1）的前提下，码本容量小，这样地址码数目少，编码的长度可最小；
- (3) 搜索码本（即检查码本，找出最接近的信号）的时间最短，这意味着处理时间短，延迟小。

该方案是基于现行预测编码的改进型，即采用码激励的矢量码表替代简单的浊音的准周期脉冲产生器，可实现变速率的混合编码。

北美 CDMA 数字蜂窝标准（IS-95）中，语音编码采用的是 QCELP（Qualcomm 码激励线性预测）语音编码技术。

QCELP 中的参量分为 3 类，即矢量码表参量、音调参量和线性预测系数参量，需要每帧更新。QCELP 采用 3 类滤波器代替典型 LPC 人工合成的滤波器，目的是改善合成语音的质量，特别是改善语音自然度。这三类滤波器包括：动态音调合成滤波器、线性预测编码滤波器和自适应共振峰合成滤波器。

QCELP 利用语音激活检测（VAD）技术，在语音激活期内，可根据不同的信噪比分别选择 4 种速率，即 8 kb/s、4 kb/s、2 kb/s 和 1 kb/s，并称它们为全速率（1）、半速率（1/2）、四分之一速率（1/4）、八分之一速率（1/8）。采用可变速率，可以使平均速率比最高速率下降一半以上。

QCELP 方案的编码器原理如图 3-24 所示。图 3-24 中， $L$  表示最佳音调滞后， $b$  表示音调滞后。

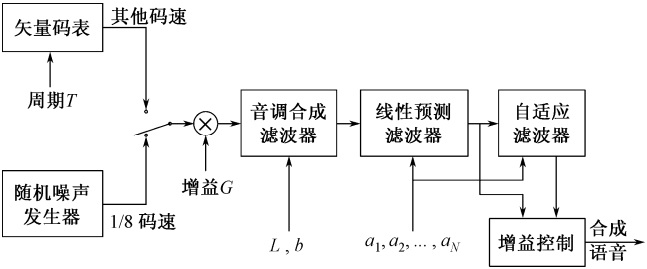


图 3-24 QCELP 方案的编码器原理图

首先对输入的模拟语音信号采取 8 kHz 取样，再按 20 ms 划分为一个语音帧，每帧含有 160 个样点，接着将 160 个样值生成 3 个参数子帧。需要注意的是，为了进一步降低 LPC 参数的编码速率，提高稳定性，需要将 LPC 参数变换为线性频谱对 LSP 参数，且每个子帧 LSP 参数是通过相邻子帧线性内插求得的，它对任何速率都是 20 ms 更新一次。

对于每个 20 ms 帧，码表中的一个矢量被选定且被量化，形成码表对数字子帧。同时生成音调特性参数子帧和线性预测滤波器参数子帧，3 种参数不断被更新，更新后的参数按一定的帧结构打包送到接收端。接收端的数据经解包，得结构参数，并从这些参数重组发送信号。

QCELP 可实现变速率语音编码。速率的确定是根据语音帧（20 ms）能量与 3 个门限的比较。每帧语音能量由语音自相关函数决定，而 3 个门限由前一帧语音自相关函数与前一帧噪声电平决定，每帧更新一次。当语音帧自相关函数大于 3 个门限时，选择全速率（9.6 kb/s）；当大于二个门限时，选择半速率（4.8 kb/s）；当仅大于一个门限时，选择 1/4 速率（2.4 kb/s）；当小于所有三个门限时，选择 1/8 速率（1.2 kb/s）；当不讲话时，用 1.2 kb/s 速率，只传背景噪声。

3.5 语音及音频压缩编码标准

语音，语言的声音，即人们说话时发出的声音，一般频率为 100~3 600 Hz。  
音频，人耳能认知和鉴别的声响，即话声、歌声或乐器声等，一般频率为 20~22 000 Hz。  
对音频信号以 11.025 kHz 抽样速率时为语音效果、22.05 kHz 抽样速率时为音乐效果、44.1 kHz 抽样速率时为高保真效果。

语音及音频编码的国际标准是语音及音频编码技术的法规和准则，语音及音频编码的地区标准也是语音及音频编码技术的重要参考和依据。了解语音及音频编码的国际标准和地区标准及其实现，对于语音及音频压缩编码技术的学习、研究具有十分重要的意义。

3.5.1 语音压缩编码标准

参与制定语音及音频编码标准的主要国际标准化组织有：国际电信联盟（ITU）、国际标准化组织（ISO）、国际电工技术委员会（IEC）、国际无线电咨询委员会（CCIR）、电子电气工程师协会（IEEE）、国际电报电话咨询委员会（CCITT）等。其中 CCITT 现在已经并入国际电信联盟（ITU）的电信标准化部（ITU-T）。

主要的地区标准化组织有：美国国家标准学会（American National Standards Institute，ANSI）、欧洲电信标准协会（European Telecommunication Standard Institute，ETSI）、日本技术咨询中心（Technology Consultancy Centre，TCC）、中国的电信标准化组织的数字音视频编解码技术标准工作组（简称 AVS 工作组）等。

下面介绍一些主要常见的国际标准和地区标准。

1. CCITT（ITU）已经公布的电话带宽语音编码标准（表 3-6）

表 3-6 CCITT（ITU）已经公布的电话带宽语音编码标准

| 标准（建议）      | CCITT | CCITT                   | ITU-T      |
|-------------|-------|-------------------------|------------|
|             | G.711 | G.721 G.723 G.726 G.727 | G.728      |
| 公布日期        | 1972  | 1984 1986 1988 1990     | 1992, 1994 |
| 编码速率/（kb/s） | 64    | —                       | 16         |
| 语音质量        | 长话质量  | 长话质量                    | 长话质量       |
| 时间复杂度/MIPS  | <1    | 1.25                    | 30         |
| 空间复杂度（RAM）  | 1 B   | <50 B                   | 2 KB       |
| 帧长度/ms      | 0.125 | 0.125                   | 0.625      |
| 编码算法类型      | PCM   | ADPCM                   | LD-CELP    |

2. ITU-T 近十几年来公布的电话带宽语音编码标准

表 3-7 ITU-T 近十几年来公布的电话带宽语音编码标准（表 3-7）

| 标准（建议）      | ITU-T G.729 | ITU-T G.723.1 | ITU-T（4 kb/s） |
|-------------|-------------|---------------|---------------|
| 公布日期        | 1996        | 1995          | 1998          |
| 编码速率/（kb/s） | 8           | 5.3, 6.3      | 4             |

续表

|            |       |            |      |
|------------|-------|------------|------|
| 语音质量       | 长话质量  | <长话质量      | 长话质量 |
| 时间复杂度/MIPS | 25    | 20         | —    |
| 空间复杂度（RAM） | 3 KB  | 2 KB       | —    |
| 帧长度/ms     | 10    | 30         | 20   |
| 编码算法类型     | CELPC | 多脉冲， CELPC | —    |

3. 北美数字蜂窝移动通信语言编码标准（表 3-8）

表 3-8 北美数字蜂窝移动通信语言编码标准

|             |           |               |            |
|-------------|-----------|---------------|------------|
| 标准          | TIA IS-54 | TIA IS-96     | TIA EVRC   |
| 公布日期        | 1989      | 1998          | 1995       |
| 编码速率/（kb/s） | 7.95      | 8.5， 4.2， 0.8 | 与 IS-96 相同 |
| 语音质量        | =RPE-LPT  | <IS-54        | <IS-54     |
| 时间复杂度/MIPS  | 20        | 20            | —          |
| 空间复杂度（RAM）  | 2 KB      | 2 KB          | —          |
| 帧长度/ms      | 20        | 20            | 20         |
| 编码算法类型      | CELPC     | CELPC         | —          |

4. 欧洲数字蜂窝移动通信语音编码标准（表 3-9）

表 3-9 欧洲数字蜂窝移动通信语音编码标准

|             |              |                 |
|-------------|--------------|-----------------|
| 标准          | ETSI-GSM 全速率 | ETSI-TCH-HS 半速率 |
| 公布日期        | 1987         | 1994            |
| 编码速率/（kb/s） | 13           | 5.6             |
| 语音质量        | <长话质量        | <长话质量           |
| 时间复杂度/MIPS  | 4.5          | 30              |
| 空间复杂度（RAM）  | 1 KB         | 4 KB            |
| 帧长度/ms      | 20           | 20              |
| 编码算法类型      | RPE-LPT      | CELPC           |

3.5.2 音频压缩编码标准

1. MPEG-1 音频压缩编码标准

国际标准化组织和国际电子技术委员会组织运动图像专家组（MPEG）于 1992 年 11 月制定的关于视频和音频信号压缩的国际标准 ISO/IEC11172,即 MPEG-1 标准.这个标准有系统、视频和音频三大主要部分。

MPEG-1 音频标准是国际上第一个高保真立体声音频编码标准。通过对 14 种音频编码方案的比较测试,最后选定了以 MUSICAM(Masking Pattern Universal Subband Integrated Coding And Multiplexing) 为基础的三层编码结构，分别称为第 I、II、III层，每一层都有不同的应用。根据不同的应用要求，使用不同的层来构成其音频编码器。第 I 层是基本模式，其余两

层有更多的处理，因而在相同的听觉质量下，将会有更高的压缩效率。

第 I 层复杂度最小，它可以用来处理相对较高的数据传输速率，每声道约为 192 kb/s。

第 II 层则是在第 I 层的基础上，用较复杂的方法来处理较低的数据率，每声道大约 96~128 kb/s。

第 III 层对联合立体声进行处理。第 III 层与第 I、II 层不同，最复杂，它对最低数据传输速率进行处理，最低数据速率大约 64 kb/s。

从第 I 层到第 III 层逐渐增加的复杂度反映在对较低的数据传输速率的处理上，第 III 层可以保证音频信号的质量。通常第 II 层与第 III 层适用于广播编码，每声道 128 kb/s 是的数据速率不会损害原始的音频信号。

MPEG-1 的各层均支持采用强度编码的联合立体声编码。将左、右高频子带取样，在一个声道内相加，但是比例因子依然是左、右独立的。译码器根据比例因子形成左、有声音道的包络。在较高子带上左、右声道的频谱形状相同，但是振幅不同。

表 3-10 给出了第 I、II、III 层的性能和应用领域，表 3-11 给出了 MPEG-1 音频第 I、II 层参数的比较。

表 3-10 MPEG-1 音频第 I、II、III 层的性能和应用领域

| 音 频 层 | 应 用         | 数据速率/(kb/s) | 音 质                | 总时延/ms |
|-------|-------------|-------------|--------------------|--------|
| I     | 数字音频磁带      | 32~448      | 每声道 192 kb/s，高保真   | 20     |
| II    | 音频和视频广播     | 32~192      | 每声道 128 kb/s，接近 CD | 40     |
| III   | 低速信道的 CD 音频 | 64          | 每声道 64 kb/s，CD 质量  | 60     |

表 3-11 给出了 MPEG-1 音频第 I、II 层参数的比较

| 参 数     | MPEG-1 音频第 I 层 | MPEG-1 音频第 II 层 |
|---------|----------------|-----------------|
| 帧长度(取样) | 384            | 1 152           |
| 子带      | 32             | 32              |
| 子带取样    | 12             | 36              |
| 快速傅里叶变换 | 512 点          | 1 024 点         |
| 比特分配    | 每子带 4 个        | 2~4 个(由子带确定)    |
| 比例因子    | 每子带 6 个        | 每子带 6~18 个      |
| 取样组     | 无              | 每子带 3 个         |

MPEG-1 中有关音频压缩的标准，已经成功应用在 VCD、CD-ROM、ISDN、数字音频广播以及视频游戏等领域中，它支持每声道为 32~224 kb/s 的 32 kHz、44.1 kHz 和 48 kHz 的 PCM 数据。

2. MP3 音频压缩编码标准

MP3 就是 MPEG-1 音频第 III 层的音频压缩编码算法，目前应用非常广泛。其文件可以在因特网上上传或下载，或者附加在电子邮件中发送。MP3 文件可以存储在硬盘中、记录在 CD-ROM 中或者存储在应用闪存的固态播放器中。在上述这些应用中，数据必须通过 MP3 播放器进行录音重放。

MP3 算法一般不带有加密或复制保护，其内容可以复制。许多共享软件和商业软件程序

均可将音乐文件转换成编码的 MP3 文件，然后再在 PC 机上对 MP3 文件解码。一般情况下，在硬盘上首次存储时是作为 WAV 或 AIFF 文件存储的，然后再用专门的硬件或软件压缩成为 MP3 文件。大多数编码器可以实现不同级别的压缩，比如，允许实现 28.8 kb/s、64 kb/s、112 kb/s、128 kb/s、192 kb/s 以及 320 kb/s 等不同传输速率的压缩。较高的传输速率可以用 44.1 kHz 取样频率提供立体声录音重放，而比较低的传输速率就不能提供了，比如，28.8 kb/s 的速率，只能用 16 kHz 取样频率产生单信号。MP3 能够在 96 kb/s 的比特率下提供好的编码效果。当然，MP3 与其他一些编译码器相比引入了比较长的编码延迟，但是这并不是一个重要问题，因为这和因特网固有的传输延迟相比是微不足道的。

MP3 以及其他 MPEG 音频编码一般不提供纠错编码，若要求进行纠错时，则必须提供额外的信道编码。通常这种要求在网络应用中是不常见的，但是，将 CRC 校验与 MPEG 数据一起传输，可以对大多数敏感的数据进行检错，并且可以进行错误隐藏（注意并非错误纠正），例如，将损坏的帧隐藏，并重发。

### 3. MPEG-2、MPEG-2BC、MPEG-2AAC 音频压缩编码标准

MPEG-1 音频编码方案的缺陷：MUSICAM 只能传送左、右两个声道。

MPEG-2 在其基础上扩展了低码率多声道编码，称为 MUSICAM 环绕声。该方案将声道数扩展至 5.1 个，即 3 个前声道（左 L、中 C 和右 R）、2 个环绕声道（左 LS、右 RS）和 1 个超低音声道 LFE（常称为 0.1 声道）。这样，就形成了 MPEG-2 音频编码标准 ISO/IEC13818-3，它于 1994 年公布。

MPEG-2 可以提供取样频率为 32 kHz、44.1 kHz 和 48 kHz 的多声道声音，也支持取样频率为 16 kHz、22.05 kHz 和 24 kHz 的单声道和立体声编码。MPEG-2 音频压缩编码标准包括 MPEG-1 音频压缩编码标准的第 I、II、III 层，使用相同的编码和译码原理。在许多情况下，MPEG-1 所设计的算法，也适用于 MPEG-2。MPEG-2 音频压缩的应用范围包括数字 HDTV 电视节目的发送以及从因特网上的下载等。

MPEG-2 向后兼容 MPEG-1，通常用 BC 表示向后兼容，则可以表示为 MPEG-2BC。多声道的 MPEG-2 音频向后兼容 MPEG-1，MPEG-2 的译码器可以接收 MPEG-1 的比特流。MPEG-1 的译码器可以从 MPEG-2 的比特流中得到立体声。

MPEG-2 AAC（高级音频编码）是运动图像专家组于 1997 年制定的 ISO/IEC13818-7 标准。它在每个声道以 64 kb/s 的比特率对立体声或多声道声音编码，也提供 5.1 声道的编码。

MPEG-2 AAC 编码不向后兼容 MPEG-1。除去兼容性的限制之外，其他性能均比 MPEG-2BC 优越。

MPEG-2AAC 支持 32 kHz、44.1 kHz 和 48 kHz 取样速率，也支持其他 8~96 kHz 的取样频率，产生的最大比特速率分别为 48 kb/s 和 576 kb/s。其输入声道的配置为 1/0（单声道）、2/0（双声道立体声）以及最高到 3/2+1 的不同多声道配置，最多可以提供 48 声道。它也支持向下混合。为了改进误差性能，其系统设计得能够在噪声存在时保持比特流同步，从而能够很好地进行噪声抵消。

MPEG-2 音频编码能传送多路音频，并能确保比特流与 MPEG-1 前向和后向兼容。由于多通道音频系统可用于卫星或陆地的电视广播，数字化音频广播以及其他诸如 CATV、视频会议、HTT（家庭电视剧场）等多媒体系统，故 MPEG-2 音频编码系统有广泛的应用，对制造商和使用者都非常有吸引力。

## 4. MPEG-4 音频压缩编码标准

虽然 MPEG-4 和其他 MPEG 保持兼容性，但是它与其他 MPEG 的编码方法很不相同。MPEG-4 是针对会话型视听系统的编码算法，其突出的特点在于具有非常低的比特率，可以在因特网及其他网络上进行操作。MPEG-4 说明了如何将真实的与合成的（计算机生成的）音频和视频形成一个对象，然后又如何组合形成完整的场景。例如，一个场景是由一个有固定背景的静态的图、人物、人的声音、音乐、插入的图形以及运动的文字所组成的。应用 MPEG-4 可以使这 6 个对象作为复用的数据流和一个场景描述一起传送。在接收端对数据流解复用，就可以将这 6 个对象分开，再按场景描述组合起来，展现给用户。

MPEG-1 和 MPEG-2 仅描述了基于帧的、最小交互能力的视频和音频的压缩、传输、存储以及处理，而 MPEG-4 则提供了在单个数据对象上的控制以及与它们相关联的方法，综合了许多不同形式的数据。不过，MPEG-4 没有说明传输机制，这样一来，就可以使用不同的方法，例如 MPEG-2 传输流、异步传输模式（ATM）和因特网上的实时传输协议（RTP）访问网络和其他数据。MPEG-4 的应用包括因特网多媒体、交互式游戏、电视会议以及视频电话等这样一类人与人之间的通信、使用光盘的交互式存储媒体、多媒体投递、网络数据库业务、HDTV 上的联合广播、远程急救系统、远程视频监控、无线电多媒体以及广播应用。

MPEG-4 音频压缩编码将高质量的音乐编码、语音编码、语音合成和计算机音乐归入一个统一的框架。MPEG-4 在现有的 MPEG 音频编码标准基础上建立，其范围从低复杂度的移动通信接入，一直到高质量的声音系统。

MPEG-4 用 MPEG-2AAC 和 MPEG-4 音频编码来支持高质量单声道、立体声和多声道信号的编码。MPEG-4 还特别针对非常低的比特率，例如针对  $64 \text{ b/s} \sim 2 \text{ kb/s}$  的自然音频进行编码。当其使用变速率编码时，甚至可以对低于  $2 \text{ kb/s}$  以下的自然音频编码，例如，可以对  $1.2 \text{ kb/s}$  速率的音频编码。

MPEG-4 也支持中等质量的音频编码。对于这一质量等级的音频信号，从使用  $8 \text{ kHz}$  的取样频率开始。

MPEG-4 支持宽带语音编码、窄带语音编码、智能语音编码、语音合成以及音频合成。MPEG-4 定义了 4 个音频的分布图，提供对极低比特率语音进行参数编码的合音矢量激活编码（HVXC）器、对窄带/宽带语音进行编码的 CELPC 编码器和一个文字到语音的接口。

## 5. AC-3 系统

AC-3 系统是 Dolby 公司开发的新一代高保真立体声音频编码系统，目的是为美国的全数字式高清晰度电视（HDTV）提供高质量的伴音。1993 年 11 月，美国高级电视系统委员会（ATSC）正式批准其大联盟高清晰度电视（GA-HDTV）系统采用 AC-3 音频编码方案。

AC-3 系统继承了 AC-2 系统的许多优点，例如，变换编码、自适应量化和比特分配、人耳心理听觉特性等，并采用了一些新技术，如指数编码、混合前/后向自适应比特分配和耦合技术等。

### 3.5.3 音频压缩文件的常见格式

（1）\*.WAV：由 Microsoft 公司和 IBM 开发，被 Windows 平台及其应用程序所支持的一种数字声音的标准声音文件格式，音质保持很好，但其数据量则会很大，因此一般不适于网

络传输或播放。

(2) \*.MP3: MP3 是 MPEG 的音频层, MPEG 音频编码具有很高的压缩比, 但 MPEG 音频文件的压缩是一种有损压缩, 存储容量小, 音质还原较好, 所以成为目前最为流行的音频格式文件。

(3) \*.RA、\*.RM 和 \*.RAM: 这些文件格式是 Real 文件的主要格式, 可以随网络带宽的不同而改变声音的质量, 在保证大多数人听到流畅声音的前提下, 令带宽较充裕的听众获得较好的音质。

(4) \*.AIF、\*.AIFF 和 \*.AU: AIF、AIFF 是 Apple 公司开发, 被 Macintosh 平台及其应用程序所支持; AU 是由 Sun Microsystems 公司与 Apple 公司为 UNIX 系统共同开发的一种经过压缩的数字声音文件格式。

(5) \*.ASF、\*.ASX 和 \*.WMA 和 \*.WAX: ASF 和 WMA 是 Microsoft 公司针对 Real 公司开发的新一代网上流式数字音频压缩技术。这种压缩技术同时兼顾了保真度和网络传输的需求, 具有一定的先进性。

(6) \*.VQF: 是 YAMAHA 公司的专用音频格式。采用减少数据流量但保持音质的方法来达到更高的压缩比。

(7) CD-DA: “CD-DA” 是数字音频光盘 Compact Disc Digital Audio 的英文缩写, 这种音乐格式在日常生活中十分常见, 它的数字化音频效果完全能够再现原始的声效且文件小。

## 3.6 图像的抽样量化与压缩编码

### 3.6.1 概述 (图像信号的特征)

图像是指景物在某种介质上的再现。图像信息, 特别是带声音的活动图像信息, 在人类感觉中起着重要作用。人类的感覺有视觉、听觉、味觉和嗅觉, 人通过五官接收和传递各种信息。其中视觉在人的五官中是最敏感的, 它约占日常生活中人们察觉信息的 60 % ~ 80 %, 它是人们所获得的最重要的信息形式。为了较好地掌握图像编码技术, 首先介绍图像信号的分类、特点与特征。

#### 1. 图像信息的分类

图像按照其内容的运动状态, 可划分为静止图像和活动图像两大类。

静止图像包括黑白二值图像 (文字、符号、真迹、图形、图书、报刊等)、黑白或彩色照片 (人物像、风景像、X 光片、工业、科技摄影灯)、高分辨率照片 (航空摄影照片、气象卫星图、资源卫星遥感照片等)。

活动图像是对运动景物连续摄取的图像, 如电影、电视中的图像。

图像也可按其色彩成分划分为黑白的、彩色的和多光谱, 还可按照其空间的性质划分为平面图像及立体图像。

#### 2. 图像信号的特点

与语音信号相比图像信号具有以下特点:

(1) 图像信号是二维以上的多维信号。例如, 在各种活动图像中, 比较简单的是黑白静

止图像，它可看做是由许多亮度不同的点组成的。

黑白静止图像在任一点的亮度是图像所在平面上直角坐标  $x$ 、 $y$  两个变量的函数，可用一个二维函数  $B(x, y)$  来描述。而黑白活动图像（如黑白电视机），由于其每一点的亮度又是随时间变化的，因而可表示为  $x$ 、 $y$ 、 $t$  的函数  $B(x, y, t)$ 。

根据三色学说，彩色图像可以用红（R）、绿（G）和蓝（B）3 个分量表示，它由 3 种颜色合成，对于彩色静止图像，由函数  $B(x, y, v)$  描述， $v$  示不同颜色。彩色活动图像可用  $B(x, y, t, v)$  表示，如果是彩色立体图像，则需用  $B(x, y, z, t, v)$  来描述。

与此相反，语音信息只随时间  $t$  变换而运动，它可以用函数  $x(t)$  表示，是一个一维信号。

（2）频带很宽，约为声音信号频谱的 1 000 倍。在数字传输体制中，如果对图像只用简单的脉码调制方式进行编码，则 1 路模拟彩色电视信号所需传码速度高达 108 MB/s。

（3）图像信号冗余度大。图像信号是高度相关的，一幅图像内部以及视频序列中相邻图像之间有大量的冗余信息。如果能有效地去除这些冗余，就可以达到压缩编码，从而减小传输信号的频带，也可以利用最少的时间和最小的空间，传输和保存图像的数据。

### 3. 图像信号的冗余

图像信号的冗余有 6 种。

（1）空间冗余。例如，图像中某区域的颜色是相同的（相关的）。

（2）时间冗余。例如，图像是静止或运动较慢时相邻两帧图像基本相同，即两幅图像有较强的时间相关性。

（3）信息熵冗余（编码冗余）。由信息论的有关原理可知，为表示图像的一个像素点，只要按其信息熵的大小分配相应比特数即可。然而对于实际图像，很难得到每个像素的信息熵，因此用相同的比特数表示每个像素，这样必然存在冗余。

信息熵冗余、空间冗余和时间冗余统称为统计冗余，因为他们都取决于图像的统计特性。

（4）结构冗余。例如，大张的草席网格，其网格的结构完全一样，图像传输时将存在结构冗余。

（5）知识冗余。有些图像信息与某些先验的基础知识有关，例如，在一般的人脸图像中，头、眼、鼻和嘴的相互位置等信息就是一些常识。此类即为知识冗余。

（6）视觉冗余。人的视觉系统并不是对于图像场的任何变化都能察觉的，如对色差信号的变化不敏感。这就可以进一步进行数据压缩和量化，使图像发生变化，只要这种变化值不超过视觉可见阈值，就认为是足够好。这就是视觉冗余。

### 4. 图像质量评价

图像质量评价是图像信息工程的基础技术之一。例如在图像通信中，将被摄物体的光学图像传送到接收端，再生出可接收的图像，其中要经过光电变换、传输、处理、记录及对其变换等过程。例如，在图像处理中记录的图像编码技术，就是在保持被编码图像一定的质量前提下，以尽可能少的码字来表示图像，以便节省信道和存储器容量。而图像增强就是为了改善图像的主观视觉显示质量，使恢复后的图像尽可能地接近原始图像质量。这些都要求有一个合理的图像质量评价方法。

图像质量的含义包括两个方面，一个是图像的逼真度，另一个是图像的可懂度。图像的逼真度是描述被评价图像与标准图像的偏离程度。图像的可懂度则是表示图像能向人或机器



提供信息的能力。

多年来，人们总是希望能够找到图像逼真度和可懂度的定量测量方法，以作为评价图像和设计图像系统的依据。目前对人的视觉系统性质还没有充分理解，对人的心理因素还找不出定量的描述方法。现在往往采用所谓主观评价方法。但这种评价方法受观察者、图像和环境影响较大，一般用 20 多人测评统计作为其主观评价价值。

### 3.6.2 图像信号的抽样与量化

目前，通信正朝着数字化等方向发展，因此，图像信号数字化具有重要意义。图像信号数字化过程包括图像信号抽样、量化和压缩编码等。下面就首先讨论图像信号抽样与量化。

#### 1. 图像信号的抽样的概念

平面的线上使连续变化的语音在时间上离散化，称为语音信号抽样。空间的面上使连续变化的图像在时间上离散化，称为图像信号抽样。

从原理上讲，图像信号的抽样和语音信号的抽样是一样的。但图像信号是二维信号，所以图像信号必须在两个方向上同时满足抽样定理。

将图像看成是一个空间上连续的区域，并把该区域划分成许多小格子，称其为样点（或像素、像元），每个样点都连续变化，对这连续变化的像素在时间上离散化，这一过程成为抽样。图像信号抽样即空间和时间上离散化。

#### 2. 抽样的频率

根据抽样定理，图像信号的抽样频率是由图像信号的上限频率决定的。

另外，在图像处理设备中，图像信号放大器的带宽一般要大于图像信号的频率上限，而处理高频信号电路的价格要高很多。即图像处理设备器件的选择也受信号上限频率的制约。

例如，我国模拟电视图像，其最高的图像频率为  $f_{\max} \approx 5.5 \text{ MHz}$ ，则抽样频率  $f_s \geq 2f_{\max} = 2 \times 5.5 \text{ MHz}$ ，或抽样周期  $T_s \leq 0.091 \mu\text{s}$ 。

实际工作应用中，兼顾设备价格，往往使抽样周期大于  $0.091 \mu\text{s}$ 。例如，取  $T_s = 0.125 \mu\text{s}$ ， $f_s \approx 8 \text{ MHz}$ 。

从理论上讲，由于其不满足抽样定理而会产生混淆失真，但实际应用中，图像质量的降低是能接受的。

#### 3. 抽样方法与方式

方法，现指为达到某种目的而采取的途径、步骤、手段等。方式，说话或做事所采取的方法和形式，有内涵之意。

抽样的方法有两类：点阵抽样方法和正文抽样方法。

在一般情况下，图像信号抽样多采用点阵抽样方法，如图 3-25 所示。点阵抽样可以是正方格（行和列等间隔分布抽样点且呈正方格排列）顺序抽样，也可以是针对电视图像的隔行抽样，还可以是纵向抽样，即每行取一点，一场采一列。

正交抽样方法如图 3-26 所示，相邻行的抽样点呈正交分布。

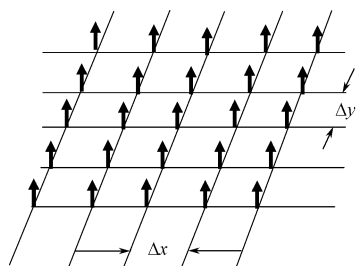


图 3-25 图像信号点阵抽样方法

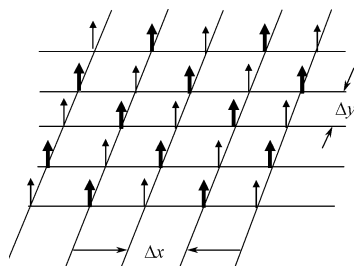


图 3-26 图像信号正交抽样方法

抽样方法的选择由图像信息的应用要求决定。对二维图像函数或者是扫描后的图像信号，可以进行等间隔的正方格顺序抽样。对变化不大的人头像为主的会议电话，统计计算分析表明，其频谱能量绝大部分集中在以原点为中心的菱形区域，而正交抽样的最大不混叠区域也呈菱形分布，所以宜采用正交抽样。对于某些图像获取速度很快（如电视摄像机），而后面的运算处理器速度跟不上的情况，就采用纵向抽样。

抽样的方式也有两类：点阵抽样方式和正文系数抽样方式。

点阵抽样方式：直接对图像的二维函数值进行抽样，即读取各离散点上的信号值，所得结果就是一个样点值阵列。

正交系数抽样方式：先将图像函数进行某种正交变换，用其变换系数作为抽样值。

实际抽样时常存在以下问题：

(1) 二维理想冲激函数作为取样函数是不可实现的，因此图像重建时，会产生边界误差的模糊现象统称抽样脉冲效应；

(2) 图像取样频率小于截止频率两倍，所以会产生混淆失真。

#### 4. 图像信号最佳量化概念

黑白图像的量化是对样点灰度级值离散化。彩色图像的量化是除样点灰度级值离散化外，颜色信号也要离散化。

将图像样点等间隔分层量化的方式称为均匀量化，不等间隔分层量化的方法称为非均匀量化。量化会产生误差，由此产生的失真即量化失真或量化噪声，对均匀量化来讲，量化分层越多，量化误差越小，但编码时占用的比特数就越多。

在一定比特数下，为减小量化误差，往往采用非均匀量化，如按图像灰度值出现概率的不同进行非均匀量化，即灰度值出现概率大的区域细量化，反之粗量化。

图像信号的量化过程与语音信号的量化过程类似。当量化器输入图像的样点幅度在判决层  $Z_k$  和  $Z_{k+1}$  之间时，量化器输出都是  $q_k$ 。所产生的量化误差分绝对误差、均方误差等数种，为使量化误差最小的量化过程或正确量化的概率大于某一值的量化称为最佳量化。

在实际图像系统中，由于成像系统引入噪声，图像本身也存在噪声，因此量化等级取得太多是没有必要的，只要量化间隔小于噪声幅度值即可。假设噪声是高斯分布，均值为 0，方差  $\sigma^2$ ，在有噪声情况下，最佳量化层的选取有两种方法：

- (1) 使正确量化的概率大于某一值；
- (2) 使量化误差的方差等于噪声方差。

此外，对非均匀分布概率密度函数还可以进行与语音压扩方法相同的量化，即先对样点

进行非线性变换，再进行均匀量化，在接收端进行相应的非线性反变换。

### 5. 黑白图像信号的矢量量化

对黑白图像序列每一个样点灰度值分别进行量化，可以称为标量量化。

为了减少量化误差，提高编码效率，也可对多个图像序列样点值进行联合量化。矢量量化就是以这种方法进行工作的。矢量量化是一种高效信源编码技术之一，基本原理是由香农在 20 世纪 50 年代推导出来的。

如将图像分成  $2 \times 2$ 、 $4 \times 4$ 、 $8 \times 8$  等小块。以  $4 \times 4$  为例，则每一小块包含 16 个像素，可把它看做一个 16 维变量，或 16 维空间中的一个矢量，各个像素的取值是该矢量的一个分量。

将前述标量量化的概念引入多维空间，就是把多维空间划分成多个胞腔（Cell），落在每个胞腔中的矢量都用同一个矢量来代表，称为码矢或码字（Cell Word），全部码字的集合称为码书（Cell Book）。

编码器的作用是判决输入的图像子块（输入矢量）落在哪个胞腔中，输出该胞腔所对应的码字在码书中的编号。解码器的作用是收到编码器送来的码字的编号，从码书上找到对应的码字，前提是解码方拥有与编码方相同的码书。

从物理意义上看，图像的矢量量化就是从图像子块的各种可能出现的情况中分组选出各自的代表，用它来代表这一组中的所有图像。

对多个图像序列点值进行联合量化，即从称为码本的码字集合中选出最适配于输入序列的一个码字，来近似一个抽样序列（一个矢量）过程。这种方法以输入序列与选出码字之间失真最小为依据。

矢量量化可归结为下列几个问题：

- （1）选取合适的失真度衡量准则，根据具体应用的要求，决定可允许的失真大小；
- （2）划分输入矢量空间及产生码书；
- （3）判决输入矢量的归属，决定输出码字在码书中的编号；
- （4）从码书中按编号查找对应的码字。

### 6. 彩色图像量化方法

目前用于色彩量化的技术主要有 3 类：

第一类是基于矢量量化的方法。由于每个像素的颜色都由红绿蓝 3 个分量组成一个三维矢量，因此可以使用矢量量化的方法对颜色量化。

第二类是用聚类的方法来产生调色板。彩色图像量化就是从原图像中挑选出少数颜色值放在一起，形成一个调色板，量化后的图像中任何一个像素的颜色值都可以从调色板中得到。这样在不影响视觉效果的情况下，大大减少原图像颜色数量，达到对原图像的颜色量化效果。

第三类算法是使用剖分技术将色彩空间分成更小子区域并从每个子区域中挑选一个典型的颜色组成调色板。

一幅真彩色图像是由红、绿、蓝三元组构成的像素阵列，每个像素中的红绿蓝分量分别代表了红色、绿色和蓝色的亮度。若每个分量的取值范围为  $0 \sim 255$ ，则一幅真彩色图像可能包含有多达  $2^{24}$  种的颜色。以第二类方法为例，色彩量化即是将相似的颜色分为一组，然后仅使用一个量化后的颜色替换它们，其过程可描述为：为源图像中的每个像素，查找或计算

用于定义目标图像中对应像素的颜色值。色彩量化技术的最终目标就是以最小的失真来减少图像的颜色数目，因此计算机图形学在色彩量化领域研究的主要内容是选择一个使量化前后图像之间差异尽可能小的最佳调色板。

常用的彩色图像的颜色量化算法是针对 RGB 彩色模式下的真彩色图像来进行的。一幅 RGB 彩色模式的真彩色图像由 3 个通道的颜色分量组成，分别是 R（红）通道的分量图像，G（绿）通道的分量图像和 B（蓝）通道的分量图像。每个通道可以应用量化单色图像的方法分别量化，可看做一幅 8 比特图像，所以一幅真彩色图像即是一幅 24 比特的彩色图像，如图 3-27 所示。彩色图像颜色量化一般要经过两个步骤：调色板设计和像素映射。

由于人眼对 R、G、B 的亮度的灵敏度不同，可以采用不同特性的压扩变换，以获得比较好的主观色彩效果。这里关键的问题是使重建和原彩色差别尽可能小。

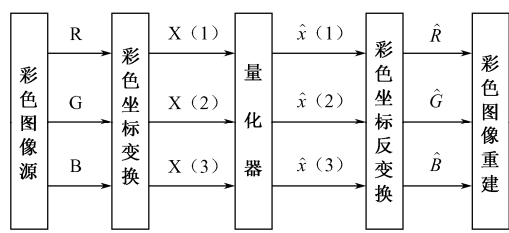


图 3-27 彩色图像量化方案

### 3.6.3 图像压缩编码基本方法

自 1948 年提出电视信号数字化后，人们就开始了图像压缩编码的研究工作，至今已有 50 多年的历史了。20 世纪 50 年代和 60 年代的图像压缩技术由于受到电路技术等制约，仅仅停留在预测编码、亚采样和内插复原等技术的研究，对视觉特性也做了一些重要的工作。1966 年，J.B.O Neal 对比分析了 DPCM 和 PCM，并提出了对电视的实验数据进行编码，1969 年进行了线性预测编码的实际实验。1969 年美国召开第一届“图像编码会议”标志着图像编码作为一门独立的学科诞生。近十年来，图像编码技术得到了迅速的发展和广泛的应用，并且臻成熟，其标志就是几个关于图像编码的国际标准的制定，即 ISO/IEC 关于静止图像的编码标准 JPEG，CCITT 关于电视电话/会议电视的视频编码标准 H.261 和 ISO/IEC 关于活动图像的编码标准 MPEG-1、MPEG-2、MPEG-4。这些标准图像编码算法融合了各种性能优良的传统图像编码方法，是对传统编码技术的总结，代表了当前图像编码的发展水平。

#### 1. 图像压缩编码分类

图像压缩编码种类很多，根据编码原理，可把编码划分为：预测编码、正交变换编码、信息熵编码、分频带（子带）编码、结构编码、矢量量化编码、基于知识编码、神经网络与小波变换编码等，其中结构编码和基于知识编码属模型编码，是宏观性的编码方法，有很高的压缩比。

根据解码后能否精确地重建原始图像，将编码分成：无损信息编码和有损信息编码。

无损编码能精确重建原始图像。利用无损压缩方法消除或减少的各种形式的冗余可以重新插入到数据中，因此，无损压缩是可逆过程，也称无失真压缩。为了消除或减少图像数据

中的冗余度，常常要用信源的统计特性或建立信源的统计模型，因此许多实用的无损压缩技术均可归结为统计编码方法。统计编码方法中常用的有 Huffman 编码、算术编码、RLE (Run Length Encoding) 编码等。此外统计编码技术在各种有损压缩方法中也有广泛的应用。

有损编码会引入失真，有损压缩法压缩了熵，信息量会减少，而损失的信息量不能再恢复，因此有损压缩是不可逆过程。有损压缩方法主要有两大类：特征提取和量化。特征提取的编码方法有模型基编码、分形编码等。量化是有损压缩最基本的形式，其优点是可以得到比无损压缩高得多的压缩比。有损压缩只能用于允许一定程度失真的情况，比如对图像、声音、视频等数据的压缩。

无损压缩和有损压缩结合形成了混合编码技术，它融合了各种不同的压缩编码技术，很多国际标准都是采用混合编码技术，如 JPEG, MPEG 等标准。利用混合编码对自然景物的灰度图像进行压缩一般可压缩几倍到十几倍，而对于自然景物的彩色图像压缩比将达到几十甚至上百倍。

## 2. 图像预测编码

图像预测编码是常用的一种压缩编码方式。它是一种针对统计冗余进行压缩的方法，可减少时间冗余和空间冗余。预测编码的特点是直观、简捷、易于实现，容易设计成实时性的硬件结构。

预测编码的基本原理是：存储以前的样本值（像素、帧），根据选取的适当模型来预测未来的样本值，对这两个样本之差进行编码。

预测编码可分为：帧内预测编码和帧间预测编码。

帧内预测编码用的最多的是 ADPCM。

帧间预测编码主要是利用活动图像序列相邻帧间的相关性，即时间冗余来达到压缩的目的。这种编码一般是针对图像块的预测编码的。它采用的技术有帧重复法、阈值法、帧内插法、运动补偿法和自适应交替帧内/帧间编码法等。

## 3. 变换域编码

变换域编码先将信号进行某种函数变换，把信号从一种描述空间变换到另一种可用较少元素表述的空间，再对变换后的信号进行编码传输，以达到降低编码传输码率的目的。

图像变换域编码通常将空间域相关的像素点通过正交变换映射到另一个频域上，使变换后的系数之间的相关性降低。在变换后的频域上满足：所有的系数相互独立、能量集中于少数几个系数上，且这些系数集中于一个最小的区域内。

常用的正交变换有傅里叶变换、沃尔什 (Walsh) 变换、离散余弦变换 (DCT)、K-L 变换、哈尔 (Haar) 变换、斜 (Slant) 变换等。

下面以离散余弦变换 (DCT) 为例，说明图像正交变换编码的原理。

例如，模拟图像经抽样后，成为平面空间上离散化的亮度值。离散化的像素点可以是  $256 \times 256$  点、 $512 \times 512$  点或  $1024 \times 768$  点，也可以是其他尺寸。

假如把整幅图像一次进行变换，则运算比较复杂，所需时间较长。通常把图像在水平方向和垂直方向分成若干子区（子平面空间），每个子区具有  $4 \times 4$ 、 $8 \times 8$  或  $16 \times 16$  点，以子区为单位进行变换，每个子区的全部像素值构成一个空间域矩阵。

设  $A$  为要编码的  $8 \times 8$  像素的图像数据矩阵， $B$  为 DCT 变换矩阵，且被定义为

$$B(i, j) = \frac{K}{2} \cos\left[\frac{i \cdot \pi \cdot (j + 0.5)}{8}\right] \quad (3-6)$$

其中： $i, j=0\sim7$  分别为行、列指数，并且当  $i=0$  时， $k=1/\sqrt{2}$ ；当  $i=1\sim7$  时， $k=1$ 。

根据  $B(i, j)$  定义，有

$$B = \begin{bmatrix} 0.3536 & 0.3536 & 0.3536 & 0.3536 & 0.3536 & 0.3536 & 0.3536 & 0.3536 \\ 0.4906 & 0.4157 & 0.2772 & 0.0975 & -0.0975 & -0.2772 & -0.4157 & -0.4906 \\ 0.4619 & 0.1913 & -0.1913 & -0.4619 & -0.4619 & -0.1913 & 0.1913 & 0.4619 \\ 0.4157 & -0.0975 & -0.4904 & -0.2778 & -0.2778 & 0.4904 & 0.0975 & -0.4157 \\ 0.3536 & -0.3536 & -0.3536 & 0.3536 & 0.3536 & -0.3536 & -0.3536 & -0.3536 \\ 0.2778 & -0.4904 & 0.0975 & 0.4157 & -0.4157 & -0.0975 & 0.4904 & -0.2778 \\ 0.1913 & -0.4619 & 0.4619 & -0.1913 & -0.1913 & 0.4619 & -0.4619 & 0.1913 \\ 0.0975 & -0.2778 & 0.4157 & -0.4904 & 0.4904 & -0.4157 & 0.2778 & -0.0975 \end{bmatrix}$$

若已采样的  $8\times8$  像素的图像数据矩阵  $A$  为

$$A = \begin{bmatrix} 150 & 170 & 132 & 185 & 147 & 190 & 215 & 220 \\ 165 & 185 & 130 & 190 & 175 & 195 & 223 & 199 \\ 155 & 163 & 180 & 220 & 202 & 173 & 197 & 170 \\ 143 & 154 & 160 & 170 & 211 & 185 & 190 & 166 \\ 130 & 140 & 172 & 190 & 193 & 150 & 180 & 140 \\ 135 & 164 & 198 & 180 & 177 & 141 & 172 & 135 \\ 170 & 190 & 163 & 140 & 165 & 132 & 160 & 140 \\ 160 & 200 & 145 & 135 & 170 & 199 & 190 & 129 \end{bmatrix}$$

并且变换系数矩阵  $C$  定义为

$$C = BAB^T$$

则有

$$C = \begin{bmatrix} 1362.2360 & -48.8294 & -39.2458 & 9.8608 & -22.6310 & 11.6491 & -63.7071 & -5.3861 \\ 62.2669 & -77.2554 & 16.1615 & -12.8255 & 35.0540 & 23.9713 & -5.6724 & -41.6991 \\ 7.5839 & -7.4009 & 75.5760 & 26.7426 & -26.9953 & -12.8395 & -10.5650 & -43.5935 \\ -18.6219 & 18.1673 & 23.2682 & -26.0828 & 17.7958 & 21.6025 & 10.0939 & 6.4744 \\ -9.1274 & -34.1931 & 0.3414 & 24.5642 & -10.6278 & -7.2710 & 7.0297 & -8.1319 \\ -9.2768 & 10.0678 & 14.2530 & -9.9202 & 10.1096 & -12.9974 & 10.0253 & 10.5422 \\ 10.7947 & 2.7326 & -29.5610 & -20.2712 & -7.3535 & 141952 & 9.6559 & 0.2945 \\ -8.3913 & 14.2379 & -8.4750 & -16.5153 & -15.8876 & 13.3316 & -2.1819 & 2.2038 \end{bmatrix}$$

其中  $C$  的大部分能量集中在左上角，意味着变换后，图像数据的相关性被去除每个系数经过量化（若为每项系数除以 30 取整）后得

$$C_q = \begin{bmatrix} 45 & -2 & -1 & 0 & -1 & 0 & -2 & 0 \\ 2 & -3 & 1 & 0 & 1 & 1 & 0 & -1 \\ 0 & 0 & 3 & 1 & -1 & 0 & 0 & -1 \\ 0 & -1 & 1 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & -1 & -1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & -1 & 1 & 0 & 0 & 0 \end{bmatrix}$$

再采用 Huffman 编码得  $C_q$  中各离散值编码为：

| 符号 | 编码    |
|----|-------|
| 0  | 0     |
| -1 | 100   |
| 1  | 101   |
| -2 | 110   |
| 2  | 11100 |
| 3  | 11101 |

|    |         |
|----|---------|
| -3 | 11110   |
| 45 | 1111100 |

结论：A 按每像素 8 bit 算，需 512 bit 传输，而  $C_c$  仅需 128 bit，压缩比为 4:1。

4. 其他图像压缩编码方法简介

1) 分形编码

分形概念是由美国数学家 B. B. Mandelbrot 在 1975 年研究不规则形状及性质时首先提出的。而利用分形作图像压缩是在 1987 年由 F.Barnsley 首先提出来的。其基本思想是寻找图像内部的自相似性，并以一定的变换来模拟，然后以变换的存储代替原图像的存储。解码时，只需将变换重复作用于初始图像，利用迭代函数系统（Iterated Function System, IFS）的理论，经过若干次迭代就可以得到原图像较好的近似。

2) 小波变换图像压缩

1989 年，S.G.Mallat 将小波变换用于多分辨率图像的描述这个过程称为小波分解。它是继傅里叶变换之后在图像压缩中的又一里程碑，它解决了很多傅里叶变换不能解决的问题。与傅里叶变换一样，小波变换的基本思想也将图像信号变为一簇基函数的加权和，而这一簇基函数是通过基本函数的平移和伸缩构成的。小波变换作为正交变换不会产生能量损失，即对小波变换后的图像进行针对性的处理才是实现压缩的关键。小波分解得到的水平和垂直细节图像的高频分量系数大部分为零，可采用非均匀量化。对系数出现概率大的用细量化，对出现概率小的系数采用粗量化。由于人眼对高频分量不敏感，可采用粗量化或舍弃。为了进一步提高压缩比，可对上述经过处理的小波系数采用游程编码或 Huffman 编码对图像进行压缩。小波变换是可逆的，通过重构算法，图像数据能够被完全恢复。

小波图像编码在较高压缩比的图像编码领域中被非常看好。在军事上，用小波编码对侦察图片进行压缩，在获得压缩比为 15 的情况下，仍可不丢失军事应用目标，能为飞机导航提供正确的军事目标位置。

3) 模型基编码

基于模型的编码技术是近几年来发展起来的一种低比特率编码方法。它利用了计算机视觉和计算机图形学中的方法和理论。其基本出发点是在编、解码两端分别建立相同的模型。基于这个模型，在编码器中用图像分析算法提取景物的参数，如形状参数、运动参数等。景物的这些参数被编码后通过信道传输到解码端，由解码器接收到的参数利用图像合成技术再重建图像。根据信源模型和编码方法的不同，模型基图像编码分为区域基编码、分割基编码、物体基编码、知识基编码和语义基编码等。

4) 神经网络

用于图像编码该研究是试图初步模仿人的视觉系统某些局部的初级功能，并将其研究成果应用到图像编码领域。将神经网络应用到图像压缩领域有两种方法：一种是直接应用，包括反向误差传播型神经网络（Back Propagation）和自组织映射（Self Organization Map）神经网络；另外一种方法是将它与经典的编码方法相结合，构成很多间接应用神经网络的编码方法。

## 3.7 图像压缩编码标准

20 世纪 80 年代以来，国际标准化组织（ISO）和国际电信联盟（ITU）组织了一批专家，开展了大量细致、全面的工作，陆续制定了一系列有关图像通信方面建议和标准，极大地推动了图像编码技术的发展与应用。这些标准可以归为两种类型：静止图像压缩标准和活动图像压缩标准（包括 ITU-T 制定的 H.263 系列和 ISO 制定的 MPEG-x 系列）。

### 3.7.1 静止图像压缩编码标准

#### 1. JBIG 标准

1988 年，ISO 和 ITU-T 成立了“联合二值图像专家组”（Joint Binary Image Expert Group, JBIG），1991 年 10 月提出了 ITU-T T.82 标准。这一标准确定了具有逐层、逐层兼容顺序和单层顺序 3 种模式的编码方法，并提出了获得任意低分辨率图像的方法。

#### 2. JPEG 标准

1986 年底，ISO 和 ITU-T 成立了联合图像专家小组（Joint Photographic Experts Group, JPEG），该小组一直致力于静止图像压缩算法的标准化工作。1991 年 3 月正式提出 ISO CD 10918 号建议草案“连续色调静止图像的数字压缩编码”（通常简称为 JPEG 标准），这是第一个适用于连续色调、多级灰度、彩色或黑白静止图像的国际标准。

JPEG 标准提供了一种无损编码的模式和 3 种有损编码模式（基于 DCT 的顺序模式、基于 DCT 的渐进模式、层次模式）。所有符合 JPEG 的编解码器都必须支持基准模式，其他模式可作为选择项，根据不同的应用目的来取舍。

尽管 JPEG 建议主要是应用于静止图像的编码技术，但是在某些场合也可将它应用于视频编辑系统。此时 JPEG 把视频序列中的每一帧当作一幅静止图像来处理，这就是所谓的 Motion JPEG 的处理方法。

#### 3. JPEG-LS 标准

JPEG 组织从 1994 年开始征集新的无损/近无损（简称 JPEG-LS）算法提案，并于 1998 年 2 月作为 ITU-T T.87（ISO/ICE 14495-1）正式公布。

#### 4. JPEG-2000 标准

JPEG-2000 标准号为 ISO/IEC 15444（ITU-TT.800），1997 年 3 月开始筹划并征集议案，2000 年 3 月、6 月、12 月，分别完成了 JPEG-2000 第一、二、五部分的 FCD，2001 年正式出版了第一部分的正式国际标准（ISO/IEC 15444-1）。JPEG-2000 规定了一系列对连续色调、二值、灰度和彩色静止图像的无失真或有失真编码方法，它抛弃了 JPEG 所采用的以 DCT 为主的区块编码方式，而采用 DWT、位平面编码和基于上下文的算术编码等一系列新技术，将图像编码的效率提高了 30 % 左右，而且提供无损和有损两种压缩方式，支持渐进传输等功能。此外，它还将 JPEG、JBIG 和 JPEGLS 统一起来，成为各种图像的通用编码方式。由于其功能强大，效率卓越，受到各界人士的广泛关注。



## 3.7.2 活动图像压缩编码标准

### 1. H.261 标准

H.261 标准是第一个基于运动补偿帧间预测和 DCT 编码的国际图像压缩编码标准,也是几十年图像编码的成果的集中体现,许多技术都被后来制定的 MPEG-1 和 MPEG-2 等编码标准所借鉴、采用和提高。它最初是作为 N-ISDN (窄带综合业务数字网) 上开展速率为  $p \times 64 \text{ kb/s}$  的双向声像业务(可视电话、视频会议)而设计的,其中  $p=1 \sim 30$ , 1990 年 12 月最终完成和批准了 H.261 标准。

H.261 标准主要使用帧间预测去除时间冗余度和 DCT 变换编码去除空间冗余度的混合编码算法。若前后两帧很相似,采用帧间方式,即进行帧间预测,并对预测误差进行二维 DCT 处理;若采用帧内方式,即对每一帧进行帧内 DCT 处理,但第一帧总是采用帧内方式。

### 2. H.263 标准

H.263 标准是 1995 年 ITU-T 为传输码率低于  $64 \text{ kb/s}$  的窄带通信信道制定的视频编码标准。该标准是在 H.261 标准的基础上进行改进扩充而发展起来的,与 H.261 标准相比,运动补偿采用半像素精度预测,由双线性内插法实现,这与 H.261 标准提出的全像素精度和循环滤波器不同。而且 H.263 标准增加了 4 种有效的压缩编码模式,分别是无限制的运动矢量模式、基于句法的算术编码模式、先进的预测模式和 PB-帧模式。

ITU-T 在 1995 年发布 H.263 标准后,根据不同的应用需要,制定了短期开发计划 H.26N 标准和长期开发计划 H.26H 标准。H.26N 标准发展成 H.263+ 标准和 H.263++ 标准。H.26L 经过 8 年时间的发展,在 2003 年作为 H.264 标准进行发布。

1998 年,ITU-T 修订发布了 H.263+ 标准。它在保证了原 H.263 标准的核心句法和语义不变的基础上,提供了 13 个新的可协商模式和其他特征,进一步提高了压缩编码性率,增强了应用的灵活性。

H.263++ 标准在 H.263+ 的基础上又增加了 6 个选项,主要是为了增强码流在恶劣信道上的抗误码性能,同时也是为了增强编码效率。

### 3. H.264 标准

H.264 标准是最新一代视频编码压缩标准,由 ITU-T/VCEG 及 ISO/IEC/MPEG 的联合视频组(JVT)研究开发,其编码效率可比现有 MPEG-4 提高 50%,是未来几年内视频编码技术研究的主要方向。2003 年 3 月公布了这一视频压缩标准的最终草案,此标准被称为 ITU-T 的 H.264 协议或 ISO/IEC 的 MPEG4 的高级视频编码部分。

H.264 的编解码框架与以前提出的标准如 H.261、H.263 及 MPEG-1/2/4 并无显著变化,也是基于混合编码的方案,新标准的性能提升在于各个部分的技术方案的改进及新算法的应用。

H.264 标准不仅是针对视频会议系统,而且涵盖了电视广播、网络流媒体、多媒体信息的数字存储、数字影院等各方面的应用。

## 4. MPEG-1 标准

MPEG (Moving Picture Experts Group) 成立于 1988 年, 该小组的主要任务是制定用于数字存储媒介中活动图像及伴音的编码标准。

MPEG-1 标准制定于 1992 年, 标准的编号为 ISO/IEC 11172, 标准的题目为“位率约为 1.5 Mb/s 时, 用于数字存储媒体的运动图像及其伴音编码设计”, 成为日后欧洲 VCD 的基础。

MPEG-1 标准包括 MPEG 视频、MPEG 音频和 MPEG 系统 3 大部分, 其中视频部分是 MPEG-1 标准的核心。MPEG-1 视频中主要使用帧内编码 (I 帧、D 帧) 和帧间编码 (P 帧、B 帧)。帧内编码包括 DCT、量化、游程编码和 Haffman 编码。帧间编码包括带运动补偿的预测法和插补法。

特别值得一提的是, MP3 是 MPEG-1 Audio Layer 3 的简称, 是压缩音频数据的国际标准, 该标准是为了提高 MPEG-1 运动画面压缩标准中的音频压缩技术而设立的。

## 5. MPEG-2 标准

由于 MPEG-1 标准的图像质量达不到电视质量, 而且不能处理隔行扫描等问题, 因此, 在 MPEG-1 的基础上, ISO 于 1994 年 11 月在新加坡会议上又通过了 MPEG-2 标准, 标准号是 IOS/IEC 13818。它的基本理念和基本结构与 MPEG-1 完全相同, 但是它允许数据率高达 100 Mb/s。

MPEG-2 标准向下兼容 MPEG-1 标准, 而 1997 年 12 月完成的 MPEG-2AAC 标准 (ISO/IEC13818-7 标准) 与 MPEG-1 标准不兼容。

MPEG-2AAC 标准对于低比特率的多声道编码能提供相当高的声音质量, 它的发展标志着标准化工作向着模块化方向演变的趋势。

## 6. MPEG-3 标准

MPEG-3 最初是为支持 HDTV 开发的压缩标准。但是, 随着 MPEG-2 标准的发展与完善, MPEG-3 还没出世就被抛弃了。

## 7. MPEG-4 标准

MPEG-4 标准是 1999 年 12 月通过的一个适应各种多媒体应用的“视听对象的编码”标准, 国际标准号是 ISO/IEC 14496。MPEG-1、MPEG-2 是将图像分割成方块处理的, 而 MPEG-4 是基于图像内容的压缩编码方法, 可以产生高压压缩比效果。它的设计思想是在超低带宽 (10 kb/s 到 1 Mb/s) 的条件下提供尽可能好的图像质量。

MPEG-4 作为第一个有交互性的动态图像标准, 更适于交互 AV (Audio/Visual) 服务以及远程监控。从目前的情况看, MPEG-4 将会在数字电视、动态图像、互联网、实时多媒体监控、移动多媒体通信、可视游戏、交互式多媒体应用等领域大显身手。

## 8. MPEG-7 标准

2001 年制定完毕的 MPEG-7 作为 MPEG 家庭中的新成员, 其正式名称为“多媒体内容描述接口” (Multimedia Content Description Interface), 它将为各种类型的多媒体信息 (包括

静态图像、图形、3D 模型、声音、语音、电视以及这些组合在一起形成的合成信息）规定一种标准化的描述。它力求能够快速且有效地检索出用户所需的不同类型的多媒体影像资料。

9. MPEG-21 标准

MPEG-21 标准的正式名称是多媒体框架，又称数字视听框架（Digital Audio-Visual Framework）。制定 MPEG-21 标准的目的是：

- （1）将不同的协议、标准、技术等有机地融合在一起；
- （2）制定新的标准；
- （3）将这些不同的标准集成在一起。MPEG-21 标准其实就是一些关键技术的集成，不同于别的 MPEG 标准的是，MPEG-21 标准将从全面的消费者需求目录开始工作，而不是从压缩或描述技术的细节开始并逐渐发展的。

当然，无论采用任何压缩技术，对最终用户来言他最为关心的几大要素无非是清晰度、存储量、稳定性、价格；而技术只有服务于用户的实际需要才能真正体现其价值，因此影响图像压缩技术未来发展的也正是上述几大要素。

在表 3-12 中列出了主要的图像压缩标准及其典型应用。

表 3-12 主要的图像压缩标准及其典型应用

| 年 份  | 俗 称       | 标 准 号                        | 适 用 信 源  | 典 型 应 用                               |
|------|-----------|------------------------------|----------|---------------------------------------|
| 1990 | P×64      | ITU-T H.261                  | 活动图像     | ISDN 上的视频会议/可视电话                      |
| 1991 | JPEG      | ITU-T T.81<br>ISO/IEC 10918  | 连续色调静止图像 | 图像库、传真、彩色印刷、数码相机等                     |
| 1991 | JPEG-1    | ITU-T T.82<br>ISO/IEC 11544  | 二值图像、图形  | G4 传真机、计算机图形、PDA 等                    |
| 1992 | MPEG-1    | ISO/IEC 11172                | 活动图像与伴音  | VCD、DAB、VOD、多媒体等                      |
| 1994 | MPEG-2    | ITU-T H.262<br>ISO/IEC 13818 | 高质量活动图像  | SVCD/DAB、VOD/MOD、多媒体游戏、DAB、DTV/HDTV 等 |
| 1995 |           | ITU-T H.263                  | 活动图像     | 视频会议/可视电话                             |
| 1998 | JPEG-LS   | ITU-T T.87<br>ISO/IEC 14495  | 连续色调静止图像 | 医学、遥感图像资料的无损/近无损                      |
| 1999 | MPEG-4    | ISO/IEC 14496                | 多媒体音像数据  | 数字电视、实时多媒体监控、互联网音像                    |
| 2001 | JPEG-2000 | ISO/IEC 15444                | 连续色调静止图像 | 图像存档、Web 浏览、数字摄像、医学成像和遥感等             |
| 2003 |           | ITU-T H.264                  | 活动图像     | 视频会议、电视广播、网络多媒体等                      |

3.8 数据信号的压缩编码

3.8.1 数据信号的压缩编码的特点

各种媒体数据，如文本、数值、图片、声音、影像等，可以保存在计算机的存储介质中，也可以将这些数据进行计算机处理或传输。

一方面，计算机的存储能力越来越大，但我们仍然常常感到存储空间不够用。因为在信

息时代，人们希望让“电脑”替自己记住尽可能多的信息，节省每一比特都有实际意义。

另一方面，在数据传输时，希望占用的信道资源越小越好，既在有限的带宽中传输更高速率的数据信号。

依据以述，所以我们要进行数据压缩，数据压缩就是通过消除数据中的冗余，达到减少数据量，缩短数据块或记录长度的过程。

数据信号的压缩编码的有它显著特点，即数据压缩必须“透明”，也就是说恢复后的文件不许有任何失真。一个符号错误就可能产生灾难性的后果，例如数据库中包含有金融交易，或者控制系统的可执行程序。压缩是在保持数据原意的前提下进行的，也称为无损压缩、无失真压缩、可逆压缩等。

典型的数据压缩方法有：Huffman 编码、算术编码、游程长度编码、Lempel-Ziv 编码等。它们都是信息熵编码或统计编码，是利用信息的相关性和概率分布的不均匀性，压缩信息冗余度，达到信息高效传输的编码方式。例，根据符号序列的统计特性，确定某种方法把符号序列变换为最短的码字序列，使各码元承载的平均信息量最大，同时又能保证无失真地恢复原来的符号序列。

### 3.8.2 数据信号的压缩编码方法

#### 1. 霍夫曼编码（Huffman 简记 HM）

霍夫曼（D. A. Huffman）于 1952 年提出一种编码方法，它完全依据字符出现概率来构造平均长度最短的异字头码字，一般就叫做霍夫曼编码。

Huffman 编码是一种不等长最佳编码方法。这里最佳是指，对一个相同概率分布的信源而言，它的平均码长比其他任何一种有效编码的都短。具体的编码方法是，对于出现概率大的信息符号，编以短字长的码，对于出现概率小的信息符号，编以长字长的码。

实现上述的编码步骤如下：

- （1）将信源符号按出现概率减小的顺序排列；
- （2）将两个最小的概率组合相加，并继续这一步骤，始终将较高的概率分支放在上部，直到概率达到 1 为止；
- （3）对每对组合中的上边一个指定为 1，下边一个指定为 0（或相反：对上边一个指定为 0，下边一个指定为 1）；
- （4）画出由概率 1 处到每个信源符号概率的路径，顺序记下沿路径的 1 和 0，所得即为该符号的霍夫曼码字。

举例：对一个 7 符号信源  $A = \{a_1, a_2, \dots, a_7\}$  做出的霍夫曼编码如图 3-28 所示。可以看出，这个编码过程实际上就是构成一棵二叉树的过程，码字都是从“树根”出发排列的。显然，编出来的码字都是异字头码，这就保证了码的惟一可译性。同时，每一个信源集合都严格按概率大小排列，编码是最佳的。

事实上，可以算出本例的信源熵为 2.61 bit，而编成码的平均字长为 2.72 bit，其编码效率为

$$\eta = \frac{H(x)}{l} = \frac{2.61}{2.72} = 96\%$$

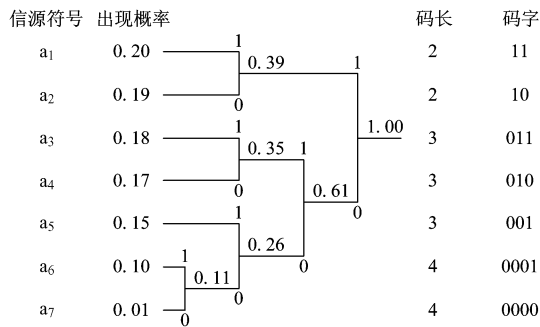


图 3-28 霍夫曼编码举例

应该指出的是：

(1) 如果概率最小的信源符号或者概率最小的中间合并结果超过两个，则继续组合相加的次序可有多种选择，故由上述过程编出的最佳码不唯一，但其平均码长相等，故不影响编码效率与数据压缩性能。另外，由于“0”与“1”的指定也是任意的，因此，同等长度的码字也不是唯一。

(2) 对不同信源概率分布，其编码效率不同，只有当信源概率分布很不均匀时，用 Huffman 方法才会收到显著的效果。

(3) 实用时，由于各符号的码长不等，霍夫曼码需要设置一定容量的缓冲寄存器（简称缓存），以便在输入速率和输出速率之间进行适配。随着微电子与计算技术的发展，霍夫曼编码已集成到许多单片的芯片中，并成为许多国际标准中的主要技术内核之一。实现了用较低的处理代价，来换取昂贵的通信开销。

## 2. 游程编码或行程编码

游程是序列中连续出现同种元素的统称。游程编码主要适用于各种连续重复字符多的场合。

游程编码可以分为定长和变长游程编码两种方式。定长游程时，游程长度的最大值固定，若连续相同的个数超过了这个固定的最大值，则超过部分进行下一轮游程编码；变长游程编码时，对不同范围的游程用不同的游程长度编码，这时需要增加标志位来说明游程长度。

游程编码的基本原理如下。

用一个特殊字符组来代替数据序列中每个长游程。这个特殊字符组一般由 3 部分组成而成：

- 第 1 部分：标号——压缩标志，压缩标志用  $S_c$  特殊标号表示，其后面使用压缩；
- 第 2 部分：字符——表示要压缩的对象（字符）；
- 第 3 部分：数字——表示压缩字符的长度。

例如，对于序列 `aaaaaaabbbcdceffff`，进行游程编码后为： $S_c a 7 S_c b 3 c d e S_c f 5$ 。

当然，实际中第 1 部分和第 3 部如何选定，要根据具体压缩的数据源的游程统计特性，以及实际需要来确定，这里仅是示意而已。游程编码对差错很敏感，要求把差错控制在一行之内。

因图像相邻像素的相关性大，游程编码在图像信号数字化压缩编码中应用较多。例如，

把图像的灰度值  $f(x, y)$ ，映射为整数对，用  $(G_i, L_i)$  表示。其中  $G_i$  表示灰度值， $L_i$  表示具有  $G_i$  灰度值的连续像素的点数，称之为行程长度。

3. 算术编码

与霍夫曼编码和游程编码不同，算术编码跳出了分组编码的范畴：从全序列出发，采用递推形式的连续编码。它不是将单个的信源符号映射成一个码字，而是将整个输入符号序列映射为实数轴上  $[0, 1]$  区间内的一个小区间，其长度等于该序列的概率；再在该小区间内选择一个代表性的二进制小数，作为实际的编码输出，从而达到了高效编码的目的。不论是否二元信源，也不论数据的概率分布如何，其平均码长均能逼近信源的熵。

早在 1948 年，香农就提出将信源符号依其概率降序排序，用符号序列累积概率的二进制表示作为对信源的编码，并从理论上论证了它的优越性。1987 年 Witten 等人发表了一个实用的算术编码程序，即 CACM87（后用于 ITU-T 的 H.263 视频压缩标准）；同期 IBM 公司发表了著名的 Q 编码器（后用于 JPEG、JPEG2000 和 JBIG 图像压缩标准），从此算术编码迅速得到了广泛的注意。

算术编码原理：

设输入符号串  $s$  取自符号集

$$S = \left\{ \begin{matrix} a_1, a_2, \cdots, a_m \\ p_1, p_2, \cdots, p_m \end{matrix} \right\}$$

$s$  后跟符号  $a_i(a_i \in S)$  扩展成符号  $sa_i$ ，空号记做  $\phi$ ，只要有一个符号  $a_i$  的序列就是  $\phi a_i$ 。算术编码的迭代关系可表示如下。

(1) 码字刷新：

$$C(sa_i) = C(s) + P(a_i)A(s) \tag{3-7}$$

其中， $P(a_i) = \sum_{k=1}^{i-1} p(a_k)$  是符号的累积概率。初始条件为  $C(\phi)=0$ ， $A(\phi)=1$  和  $P(\phi)=0$ ， $p(\phi)=1$ 。

(2) 区间刷新：

$$A(sa_i) = p(a_i)A(s) \tag{3-8}$$

(3) 区间范围：

$$[ C(sa_i), \quad C(sa_i) + A(sa_i) \quad ) \tag{3-9}$$

可见，算术编码在传输任何符号  $a_i$  之前，信息的完整范围是

$$[ C(\phi), C(\phi) + A(\phi) ) = [ 0, 1 )$$

表示为  $0 \leq p(a_i) < 1$ 。

当处理  $a_i$  时，这一区间的宽度  $A(s)$  就依据  $a_i$  的出现概率  $p(a_i)$  而变窄。

符号序列越长，相应的子区间的宽度就越窄，编码表示该区间所需的位数也越多。而大概率符号比小概率符号使区间缩窄的范围要小，所增加的编码位数也少。另外从上述迭代公式可知，符号串每一步新扩展的码字  $C(sa_i)$  都是由原符号串的码字  $C(s)$  与新区间宽度  $A(sa_i)$  的算术相加而得，“算术码”一词便由此而来。

例如，设某信源取自符号集  $S = (a, b, c, d, e, !)$ ，其中前 5 个符号为实际英文字母，而最后一个符号“!”，则用来表示编码结束，各符号概率和初始子区间范围  $[P(a_{i-1}), a_i)$  如表 3-13 所示。

表 3-13 各符号概率和初始子区间范围

| 字 符 | 概 率 | 累 计 概 率 | 区 间 范 围    |
|-----|-----|---------|------------|
| a   | 0.2 | 0       | [0, 0.2)   |
| b   | 0.1 | 0.2     | [0.2, 0.3) |
| c   | 0.1 | 0.3     | [0.3, 0.4) |
| d   | 0.3 | 0.4     | [0.4, 0.7) |
| e   | 0.2 | 0.7     | [0.7, 0.9) |
| !   | 0.1 | 0.9     | [0.9, 1.0) |

设待编码的字符串为单词 “dead!”，编码器和解码器都知道区间初值为[0，1)。在编码第 1 个字母 “d” 时

$$P(d)=\sum_{k=1}^{i-1}p(a_k)=P(d)=p(a)+p(b)+p(c)=0.2+0.1+0.1=0.4$$
$$C(sd)=C(\phi d)=C(\phi)+P(d)A(\phi)=0+0.4\times1=0.4$$
$$A(sd)=A(\phi d)=p(d)A(\phi)=0.3\times1=0.3$$

“d” 编码完成后，区间范围由[0，1) 变为  
[ C(sd), C(sd)+ A(sd) )=[ 0.4, 0.7 )

在编码第 2 个字母 “e” 时

$$P(e)=\sum_{k=1}^{i-1}p(a_k)=P(e)=p(a)+p(b)+p(c)+p(d)=0.2+0.1+0.1+0.3=0.7$$
$$C(se)=C(sd)+P(e)A(sd)=0.4+0.7\times0.3=0.61$$
$$A(se)=p(e)A(sd)=0.2\times0.3=0.06$$

“e” 编码完成后，区间范围由[0，1) 变为  
[ C(se), C(se)+ A(se) )=[ 0.61, 0.67 )

依此类推，可将这一过程表示为表 3-14。

表 3-14 6 元信源的算术编码过程

|        | [ C(sa <sub>i</sub> ), C(sa <sub>i</sub> ) + A(sa <sub>i</sub> ) ) | A(sa <sub>i</sub> ) |
|--------|--|---------------------|
| 初始值    | [0, 0.2)   | 1.0                 |
| 编完 d 后 | [0.4, 0.7)   | 0.3                 |
| 编完 e 后 | [0.61, 0.67)   | 0.006               |
| 编完 a 后 | [0.61, 0.622)  | 0.012               |
| 编完 d 后 | [0.614 8, 0.618 4)   | 0.003 6             |
| 编完! 后  | [0.618 04, 0.618 4)  | 0.000 36            |

如果解码器也知道这一最后的区间范围[0.618 04, 0.618 4)，则它立即就可解得第 1 个字符为 d，因为在表 3-13 中，只有 d 的区间范围能包含[0.618 04, 0.618 4)。

此后，解码区间由初始值[0，1) 变为[0.4, 0.7)。而得到这一范围后再对所有字符按式 (3-7)、式 (3-8) 和式 (3-9) 计算，并与最终的区间范围[0.618 04, 0.618 4) 相比较，又不难解出第 2 个字符为 e。

依此类推，解码器就将唯一地解出字符串“dead!”。这就是算术编码和解码的基本原理。由上例我们知道：算术编码就是将每个字符串  $s$  都与一个子区间  $[C(s), C(s) + A(s))$  相对应，其中子区间宽度  $A(s) \leq 1$  是有效的编码空间，而整个算术编码过程，实际上就是依据字符的发生概率，对码区间的分割过程，即子区间宽度与正编码字符发生概率相乘的过程。

实际上解码器无须完全知道最终编码区间  $[0.618\ 04, 0.618\ 4)$  的两个端点值，知道其间的某个值已足够了，例如，0.618 1, 0.618 2, 0.618 3, 甚至 0.618 040 1 均可。但对本例若不设置一个专门的结束码“!”（或采取其他措施，如预知字符串  $s$  的长度），则如果只传输一个值 0.0，解码器将无法判断被编码的究竟是字符“a”，还是字符串“aa”，“aaa”，“...”

在固定编码模式中，概率统计与区间分配将直接影响编码效率。对于字符串“dead!”而言，概率分配“d(0.4)、e(0.2)、a(0.2)、!(0.2)”为最好。而在算术编码的自适应模式中，各符号的概率初始值都相同，但依据实际出现的符号而相应地改变。只要编、解码器使用相同的初始值和相同的修改方法，则其概率模型将保持一致：编码器对下一个字符完成编码后再改变概率模型；解码器则根据当前模式完成解码后再修改自己的概率模型。

## 思考与练习

- 3-1 说明压缩编码的分类。
- 3-2 说明带通抽样定理。
- 3-3 说明均匀量化与非均匀量化、绝对量化与相对量化的区别。
- 3-4 画出差分脉码调制（DPCM）原理组成框图，说明工作原理。
- 3-5 画出语音信号产生的数学模型，说明参量编码的基本原理。
- 3-6 简述线性预测编码（LPC）的原理，指出改进思路。
- 3-7 说明 MP3 音频压缩编码标准和 MPEG-4 音频压缩编码标准的特点。
- 3-8 说明图像信号的 6 种冗余。
- 3-9 什么是图像矢量量化与图像信号最佳量化？
- 3-10 说明图像抽样方法与方式区别。
- 3-11 说明正交变换编码的概念。
- 3-12 说明多媒体音像数据 MPEG-4 标准的概念。
- 3-13 介绍几种数据信号的压缩编码方法。



## 第4章 数据基带传输系统与技术

数字和数据基带传输系统类似，所以本章不区别数字和数据的概念。

在数据基带传输系统中，其传输对象通常是二元数字信息，它可能来自计算机、电传打字机或其他数字设备的各种数据代码，也可能来自数字电话终端的 PCM 脉冲编码信号等。

由于未经调制的电脉冲数字信号所占据的频带通常从直流或低频开始，我们常称为基带信号。

在某些有线信道中，特别是传输距离不太远的情况下，基带信号可以直接传送，我们称之为数据信号的基带传输。

在另外一些信道，特别是无线信道和光信道中，数据基带信号则必须经过调制，将信号频谱搬移到高频处才能在信道中传输，我们把这种传输称为数据信号的调制传输（或载波传输、频带传输）。

如果把无线调制与解调过程看做广义信道的一部分，则任何数据传输系统均可等效为基带传输系统，因此掌握数据信号的基带传输原理是十分重要的。

### 4.1 数据基带信号编码技术

#### 4.1.1 数据基带信号的编码原则

如前所述，数据基带信号是用数字信息的电脉冲表示，不同形式的数据基带信号（又称为码型）具有不同的频谱结构，合理地设计基带信号的频谱结构，使数据信息变换为更适合于给定信道传输，是基带传输首先要考虑的问题。

事实上，在数据设备内部用导线连接起来的各器件之间就是用一些最简单的数据基带信号来传输码元。这些最简单的数据基带信号的频谱中含有丰富的低频分量乃至直流分量。由于传输距离很近，高频分量衰减也不大。但是数据设备之间长距离有线传输时高频分量衰减随距离的增加而增大，同时数据设备中往往还存在隔直流电容或耦合变压器，因而传输频带的高频和低频部分均受限。此时必须考虑码型选择问题。

归纳起来，在设计数据基带信号码型时应考虑以下原则：

（1）码型中低频、高频分量尽量少。传输码中高、低频能量在传输中均有大的衰减，且低频时要求元件尺寸大，高频能量对邻近线路造成相互干扰（串扰、串音）。这样做还可以节省传输频带，提高信道的频谱利用率。

（2）码型中应包含定时信息，以便定时提取。在基带传输系统中，位定时信息是接收端再生原始信息所必需的。在某些应用中位定时信息可以用单独的信道与基带信号同时传输，但在远距离传输系统中这常常是不经济的。因而需要从基带信号中提取位定时信息，这就要求基带信号或经简单的非线性变换后能产生出位定时号频谱。

（3）码型变换（编译码）设备要简单可靠。

（4）码型具有一定检错能力，若传输码型有一定的规律性，则就可根据这一规律性来检

测传输质量，以便做到自动监测。

(5) 编码方案对发送消息类型不应有任何限制，适合于所有的二进制信号。这种与信源的统计特性无关的特性称为对信源具有透明性。

(6) 低误码增殖。误码增殖是指单个的数字传输错误在收端解码时，造成错误码元的平均个数增加。从传输质量要求出发，希望它越小越好。

(7) 高的编码效率。

上述各项原则并不是任何基带传输码型均能完全满足，往往是依照实际要求满足其中的若干项。

数据基带信号的码型种类繁多，根据各种数据基带信号中每个码元的幅度值不同，可以把它们归纳分类为二条码、三元码和多元码，本章不可能一一叙述。

因基带传输时接口码型是一个非常重要的参数，也是无线与有线传输的界面，所以我们仅就目前应用的一些重要码型进行介绍。

## 4.1.2 二条码

最简单的二条码中基带信号的波形为矩形，幅度取值只有两种电平。常用的有如下几种，它们的波形如图 4-1 所示。

### 1. 单极性非归零码

在这种二条码中用高电平和低电平（常为零点平）分别表示二进制信息的“1”和“0”，在整个码元期间电平保持不变，常记做 NRZ (L)，如图 4-1 (a) 所示。

### 2. 双极性非归零码

在这种二条码中用正电平和负电平分别表示“1”和“0”，与单极性非归零码相同的是整个码元期间电平保持不变，而在这种码型中部不存在零电平，如图 4-1 (b) 所示。

### 3. 单极性归零码

与单极性非归零码不同，发送“1”时在整个码元期间电平只持续一段时间，在码元的其余时间内则返回到零电平，常记做 RZ (L)，如图 4-1 (c) 所示。

不难看出还应当存在一种双极性归零码，它兼有双极性和归零的特点。但由于它的幅度取值存在 3 种电平，因此我们将它们归入三元码。

由信号分析可知，上述三种最简单的二条码存在以下问题：

(1) 其功率谱中有丰富的低频乃至直流分量（一般发送“1”和“0”概率不等）。这对于大多数采用交流耦合的有线信道来说是不允许的。

(2) 当信息中包含长串连续“1”或“0”时，非归零码呈现出连续的固定电平。由于信号中不出现跳变，因而无法提取定时信息。单极性归零码在传送连“0”时，存在同样的问题。

(3) 它们不具有检测错误的能力。由于信道频带受限并且存在其他干扰，经传输信道后基带信号波形会产生畸变，从而导致接收端错误地恢复原始信息。在上述二条码信息中每个“1”和“0”分别独立地相应于某个传输电平，相邻信号之间不存在任何制约，正是这种不相

关性使这些基带信号不具有检测错误信号状态的能力。

由于这些问题，它们通常只用于设备内或很近距离（如计算机 CPU 与外设之间的连接）的信息传递。

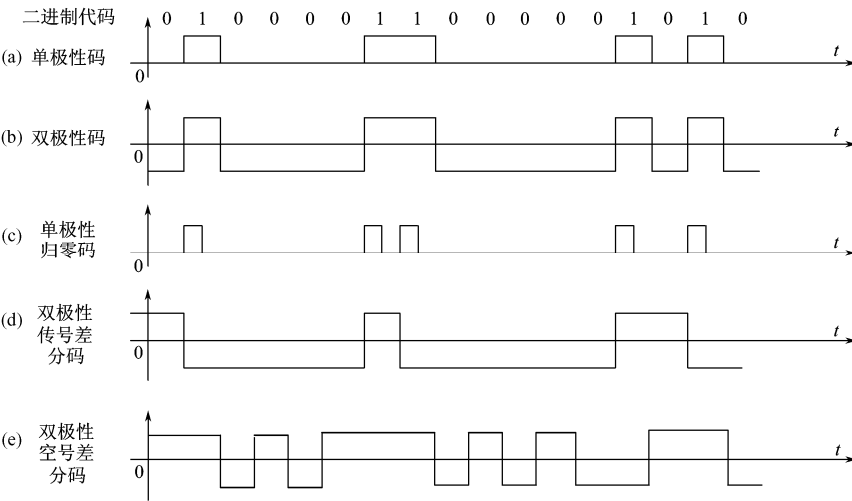


图 4-1 常用的几种二进制码波形

4. 差分码

在差分码中，“1”、“0” 分别用电平跳变或不变来表示。若用电平跳变来表示“1”，则称为传号差分码（在电报通信中常把“1”称为传号，把“0”称为空号）。

若用电平跳变来表示“0”，则称为空号差分码。

图 4-1（d）和（e）中分别画出传号差分码和空号差分码，通常分别记为 NRZ（M）和 NRZ（S）。

差分码并未解决 3 种二进制码所存在的问题。但由于它的电平与信码“1”、“0” 之间不存在绝对的对应关系，而是用电平的相对变化来传输信息，因此，它可以用来解决 PSK 相位键控解调时，因接收端本地载波相为倒置而引起的信息“1”、“0” 倒换问题，所以得到广泛的应用。由于差分码中电平具有相对意义，因而又称为相对码。

5. 数字双相码

数字双相码又称为分相码或曼彻斯特码。它用一个周期方波（方波周期内负、正电平各占一半）表示“1”，而用同周期方波的反相波形表示“0”。

数字双相码可以用单极性非归零码 NRZ（L）与定时信号的模 2 和来产生，如图 4-1（a）所示。

为了解决因极性反转而引起的译码错误，可以采用差分码的概念，将数字双相码中用绝对电平表示的波形改为用电平的相对变化来表示。为此，将差分码 NRZ（M）与定时信号模 2 和，所产生的基带信号中，相邻周期的方波如果同相则代表“0”，反相表示“1”。这种码型称为条件双相码，常记做 CDP 码，如图 4-2（a）所示。

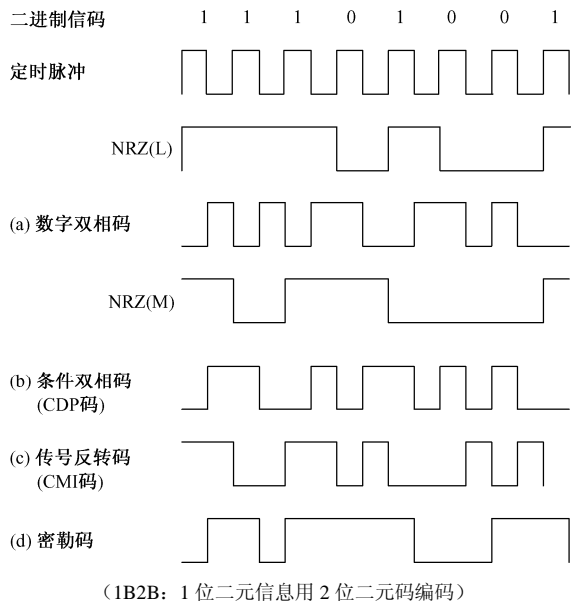


图 4-2 部分 1B2B 码波形

由于双相码在每个码元间隔的中心部分都存在电平跳变，因此在频谱中存在很强的定时分量，它不受信源统计特性的影响。

此外，由于方波周期内正、负电平各占一半，因而不存在直流分量。显然，上述优点是用频带加倍来换取的。

双相码适用于数据终端设备在短距离上的传输。如由 Xerox、DEC 和 Intel 公司共同开发的“以太”本地数据网中采用数字双相码作为线路传输码型。

## 6. 传号反转码

传号反转码（CMI 码）与数字双相码类似，也是一种二电平非归零码，如图 4-1（b）所示。CMI 码已纳入 CCITT 建议，作为脉冲编码调制 4 次群的接口码型。

在 CMI 码中，“1”（传号）交替地用正相和反相周期方波来表示，“0”（空号）用不变相周期方波表示。

CMI 码也没有直流分量，却有频繁出现的波形跳变，便于恢复定时信号。

由波形可知，用负跳变可直接提取位定时信号，不会产生相位不确定问题。

CMI 码的另一个特点是它有检测错误的能力。这是因为“1”相当于用交替的“00”和“11”两位码组表示，而“0”则固定地用“01”表示，在正常情况下“10”是不可能出现的，连续的“00”和“11”也是不可能出现的，这种相关性可以用来检测因信道而产生的部分错误。

由于 CMI 码易于实现，且具有上述优点，因此在高次群脉冲编码调制终端设备中广泛用作接口码型，在速率低于 8 448 kb/s 的光纤数字传输系统中则被推荐为线路传输码型。

在数字双相码和 CMI 码中，原始的二元信息在编码后都用一组两位的二元码来表示，因此这类码又称为 1B2B 码型。在某些文献中 1B2B 码特指 CMI 码。

7. 密勒码

密勒码又称延迟调制，它是数字双相码的一种变形。

在密勒码中，“1”用码元周期中点处出现跳变来表示，而对于“0”则有两种情况：当出现单个“0”时，在码元周期内不出现跳变；但若遇到连“0”时，则在前一个“0”结束（也就是后一个“0”开始）时出现电平跳变。

由上述编码规则可知，当两个“1”之间有一个“0”时，则在第一个“1”的码元周期中点与第二个“1”的码元周期中点之间无电平跳变，此时密勒码中出现最大宽度，即二个码元周期。

密勒码实际上是数字双相码经过一级触发器后得到的波形，如图 4-2（d）所示。因此，密勒码是数字双相码的差分形式，它也能克服数字双相码中存在的相位不确定的问题。

功率谱分析表明，密勒码的信号能量主要集中在二分之一码速以下的频率范围内，直流分量很小，频带宽度约为数字双相码的一半，其功率谱如图 4-3 所示。

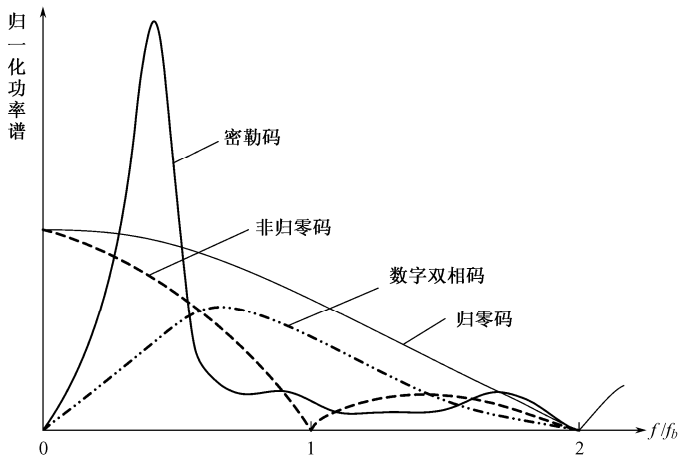


图 4-3 二元码的单边功率谱

利用密勒码最大宽度为二个码元周期，而最小宽度为一个码元周期这一特点，可以检测传输误码或线路故障。密勒码最初用于气象卫星及磁带记录，现用于传递低速数据的基带数传机。

4.1.3 三元码

在三元码数字基带信号中，信号幅度取值有 3 个：+1、0、-1。由于实现时并不是将二进制数变为三进制，而是某种特定取代，因此又称为准三元码或伪三元码。三元码种类很多，被广泛地用做脉冲编码调制的线路传输码型。

1. 传号交替反转码（AMI 码）

传号交替反转码常记做 AMI 码。在 AMI 码中，二进制信息“0”变换为三元码序列中的“0”，二进制信息“1”则交替地变换为“+1”和“-1”的归零码，通常脉冲宽度为码元周期

之半, 如图 4-4 (a) 所示。

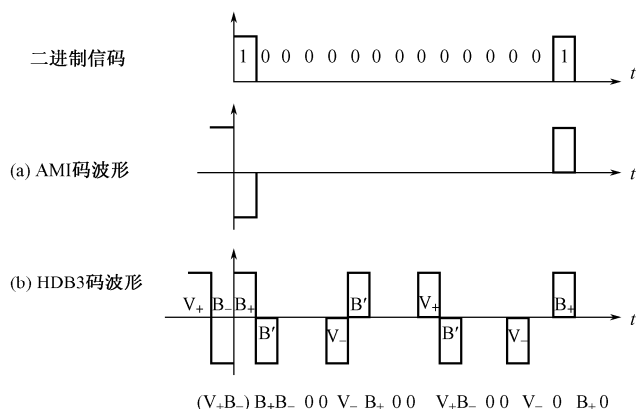


图 4-4 三元码波形

理论分析表明, AMI 码的功率谱如图 4-5 所示, 图 4-4 中还画出二元双极性非归零码的功率谱, 以资比较。AMI 码的功率谱中无直流分量, 低频分量较小, 能量集中在频率为  $1/2$  码速之处。位定时频率(即码速频率)分量虽然为 0, 但只要将基带信号经全波整流变为二元归零码, 即可得到定时信号。

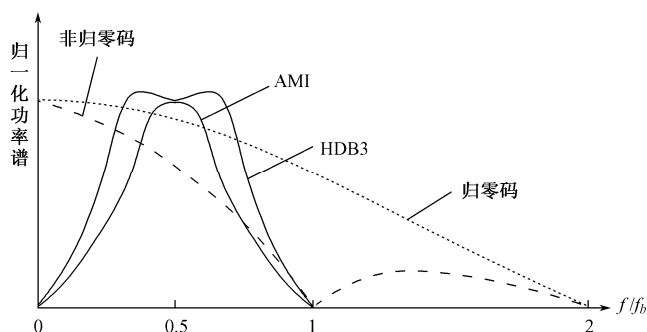


图 4-5 AMI 和 HDB3 码的功率谱

AMI 码具有检错能力，如在传输过程中因传号极性交替规律受到破坏而出现误码，则在接收端很容易发现这种错误，例如：

|        |    |   |    |   |   |   |   |   |    |   |   |    |    |
|--------|----|---|----|---|---|---|---|---|----|---|---|----|----|
| 二进制    | 1  | 0 | 1  | 0 | 0 | 0 | 0 | 0 | 1  | 0 | 0 | 1  | 1  |
| 发送AMI码 | +1 | 0 | -1 | 0 | 0 | 0 | 0 | 0 | +1 | 0 | 0 | -1 | +1 |
| 接收AMI码 | +1 | 0 | -1 | 0 | 0 |   | 0 | 0 | +1 | 0 | 0 | -1 | +1 |
|        |    |   |    |   |   |   |   |   |    |   |   |    |    |
|        |    |   |    |   |   |   |   |   |    |   |   |    |    |
|        |    |   |    |   |   |   |   |   |    |   |   |    |    |
|        |    |   |    |   |   |   |   |   |    |   |   |    |    |
|        |    |   |    |   |   |   |   |   |    |   |   |    |    |
|        |    |   |    |   |   |   |   |   |    |   |   |    |    |

↑                  ↑

破坏极性交替规律

从信息论观点看,AMI 码之所以有检错能力是因为它含有冗余的信息量。事实上,任何具有检错能力的码型必须带有这种冗余型,否则便丧失了检错能力。

AMI 码的主要缺点是它的性能与信源统计特性有密切关系。它的功率谱形状随信息中传号率（即出现“1”的概率）而变化。更要注意的是，当信息中出现连“0”码时，由于 AMI

码中常时间不出现电平跳变，因而定时提取遇到困难。

2. HDBn 码

HDBn 码是  $n$  阶高密度双极性码的缩写。在 HDBn 码中信息“1”也交替地变换为+1 与 -1 的半占空归零码，但与 AMI 码不同的是，HDBn 码中的连“0”数被限制为小于或等于  $n$ 。当信息中出现  $n+1$  个连“0”码时就用特定码组来取代，这种特定码组称为取代节。为了在接收端识别出取代节，人为地在取代节中设置“破坏点”，在这些“破坏点”处传号极性交替规律受到破坏。

HDBn 码是一种模态取代码，它有两种取代节：

$$\underbrace{B \ 0 \cdots 0}_{n+1} \ V \text{ 或 } \underbrace{0 \cdots 0}_{n+1} \ V$$

其中 B 表示符合极性交替规律的传号，V 表示破坏极性交替规律的传号，V 就是破坏点。这两种取代节的选取原则是：使任意两个相邻 V 脉冲间的 B 脉冲数目为奇数。这样，相邻 V 脉冲的极性也满足交替规律，因而整个信号仍保持无直流分量。

HDBn 码中应用最广泛的是 3 阶高密度双极性码，即 HDB3 码。在 HDB3 中每当出现 4 个连“0”码时用取代节 B00V 或 000V 代替。例如：

$$\begin{aligned} \text{二进制信息} & \quad 1 \ 0000 \ 0000 \ 0000 \ 0100 \\ \text{HDB3 码 (V}_+\text{B}_-\text{)} & \quad B \ B00V \ B00V \ B00V \ 0B00 \\ & \quad (V_+B_-) \ B_+ \ B_-00V_- \ B_+00V_+ \ B_-00V_- \ 0B_+00 \end{aligned}$$

上述 HDB3 码波形示于图 4-4 (b) 中，它是在信息序列前一破坏点为  $V_+$ ，且它至第一个连“0”串前有奇数个 B 情况下得到的。但若前一破坏点为  $V_-$ ，且它至第一个连“0”串前有偶数个 B，则 HDB3 码变为另一种形式。

以上 HDB3 码中  $B_+$ 、 $B_-$  分别表示符合极性交替规律的正脉冲和负脉冲。 $V_+$ 、 $V_-$  分别表示极性交替规律的正脉冲和负脉冲。

4.1.4 多元码

为了进一步提高频带利用率，可以采用信号幅度具有更多取值的数字基带信号，即多元码。在多元码中，每个符号可以用来表示一个二进制码组，因而成倍地提高了频带利用率。对于  $n$  位二进制码组来说，可以用  $M=2^n$  码元，进行传输。与二元气码元传输时所需要的信道频带可降为  $1/n$ ，即频带利用率提高为  $n$  倍。

由于频带利用率高，多元码在频带受限的高速数字传输系统中得到广泛应用。在综合业务数字网中，以电话线为传输媒介的数字用户环的基本传输速率为 144 kb/s。在这种频带受限的基带传输系统中，线路码型的选择是重要问题，CCITT 已将一种四元码 2B1Q 列为建议标准。

2B1Q 中，2 个二进制码元用 1 个四元码表示，如图 4-6 所示。为了减小在接收时因错误判定幅度电平而引起的误比特率，通常采用格雷码表示信息，此时相邻幅度电平所对应的码组之间只发生 1 个比特错误。

多元码不仅用于基带传输，而且更广泛地应用于调制传输中，关于这方面的问题将在第 5 章讲述。

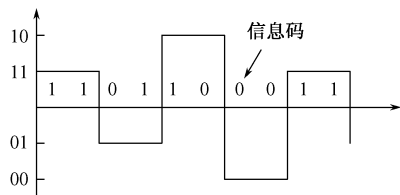


图 4-6 2B1Q 基本信号

以上介绍的几种码型，其波形均为矩形脉冲。实际上，基带传输系统中各处的信号波形可以是矩形脉冲，也可以是其他形状，如升余弦、三角形等。

## 4.2 无码间串扰的基带系统特性

### 4.2.1 基带系统组成及各点波形

数字基带系统典型模型如图 4-7 所示，为了便于分析，把数字基带信号的产生过程分成码型编码和波形形成两步：码型编码的作用是将数字信号编成合适的代码，其输出信号为脉冲序列，一般用脉冲形成器表示，波形形成网络的作用则是将每个脉冲转换为一个更适合信道传输的波形，一般用发送滤波器和接收滤波器概括。

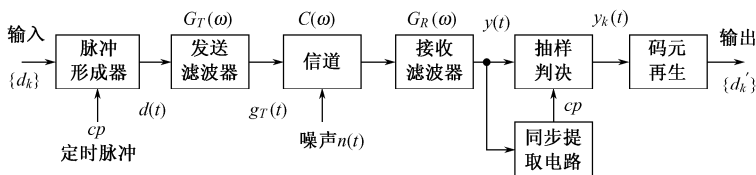


图 4-7 数字基带系统典型模型

传输信道在这里是广义的，它可以是传输介质（电缆、双绞线等），也可以是带调制解调器的调制信道。接收滤波器的作用是限制传输信道所引入的噪声，并得到所需形状的接收波形。再生判决电路将接收到的波形恢复为脉冲序列。最后经码型译码，得到发送端所要传递的原始信码。数字基带传输系统各点的波形如图 4-8 所示。

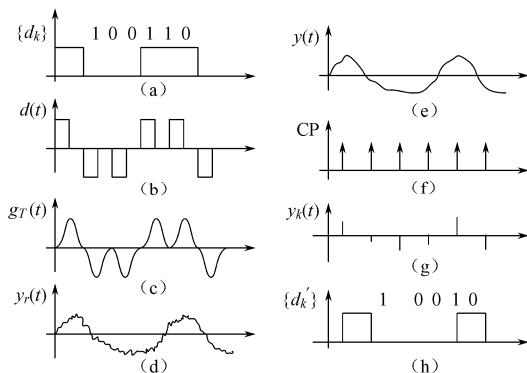


图 4-8 数字基带传输系统各点波形举例



传输过程中第 4 个码元发生了误码，产生该误码的原因就是信道加性噪声和频率特性的不理想。

### 4.2.2 基带传输特性与冲激响应

根据图 4-7 我们可以得到总的数字基带传输函数  $H(\omega)$  和冲激响应  $h(t)$  为

$$\begin{aligned} h(t) &= \frac{1}{2\pi} \int_{-\infty}^{\infty} H(\omega) e^{j\omega t} d\omega \\ &= \frac{1}{2\pi} \int_{-\infty}^{\infty} G_T(\omega) C(\omega) G_R(\omega) e^{j\omega t} d\omega \\ H(\omega) &= G_T(\omega) \cdot C(\omega) \cdot G_R(\omega) \end{aligned} \tag{4-1}$$

为简化分析，假定输入基本脉冲为理想单位冲激响应  $\delta(t)$ ，这样发送滤波器的输入信号可表示为

$$d(t) = \sum_{n=-\infty}^{\infty} a_n \delta(t - kT_b) \tag{4-2}$$

其中  $a_n$  是第  $n$  个码元，对于单极性二进制数字信号， $a_n$  的取值为 0、1，双极性二进制数字信号， $a_n$  的取值为 +1、-1。由图 4-7 可以得到

$$y(t) = \sum_{n=-\infty}^{\infty} a_n h(t - nT_b) + n_R(t) \tag{4-3}$$

式中  $n_R(t)$  是加性噪声  $n(t)$  通过接收滤波器后的滤波输出，则在第  $k$  个码元时刻  $t=(kT_b+t_0)$

$$\begin{aligned} y(kT_b + t_0) &= \sum_{n=-\infty}^{\infty} a_n h(kT_b + t_0 - nT_b) + n_R(kT_b + t_0) \\ &= \sum_{n=-\infty}^{\infty} a_n h[(k - n)T_b + t_0] + n_R(kT_b + t_0) \end{aligned}$$

若在上式中把  $n=k$  的这一项单独列出后，

$$y(kT_b + t_0) = a_k h(t_0) + \sum_{\substack{n=-\infty \\ n \neq k}}^{\infty} a_n h[(k - n)T_b + t_0] + n_R(kT_b + t_0) \tag{4-4}$$

第一项  $a_k h(t_0)$  是第  $k$  个码元本身产生的信号样值，它是判断  $a_k$  存在的依据。

第二项  $\sum_{\substack{n=-\infty \\ n \neq k}}^{\infty} a_n h[(k - n)T_b + t_0]$  是由无穷多项组成，它代表除第  $k$  个码元以外的其他码元产

生的不需要的串扰值，称为码间串扰。通常与第  $k$  个码元越近的码元，产生的串扰越大，反之串扰越小。在正常通信时我们总是希望这一项为零，但要消除码间串扰，使  $a_n$  始终恒等于零不可能，因它要传输的信号是随机的，所以只有让  $\sum_{\substack{n=-\infty \\ n \neq k}}^{\infty} h[(k - n)T_b + t_0] = 0$ ，也就是选择一个合适的传输特性  $H(\omega)$  即可，这个问题我们将在后面重点介绍。

第三项  $n_R(kT_b+t_0)$  是第  $k$  个码元抽样判决时刻噪声的瞬时值，它是由信道噪声引起，是一个随机变量，也要影响第  $k$  个码元的正确判决。我们同样希望它为零，或者越小越好。

4.2.3 系统无码间串扰的条件

在不考虑噪声干扰的条件下，对式（4-4）右边第二项进行分析，假设  $t_0=0$ ，此时无码间串扰的数字基带传输的基本准则为

$$\sum_{n=-\infty}^{\infty} h[(k-n)T_b] = \begin{cases} 1(\text{常数或规一化值}) & n = k \\ 0 & n \neq k \end{cases} \tag{4-5}$$

将式（4-5）改写为分段求和积分的形式， $\omega$ 轴分成间隔为  $2\pi/T_b$  的很多小段，如图 4-9 所示。

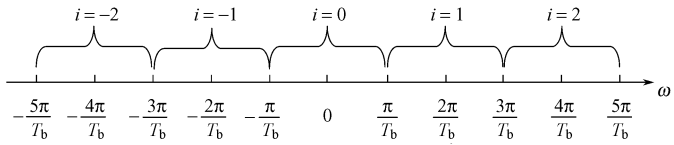


图 4-9  $\omega$ 轴分成间隔为  $2\pi/T_b$  的小段

由此得到无码间串扰的系统冲激响应和传输特性为：

$$h(kT_b) = \frac{1}{2\pi} \int_{-\pi/T_b}^{\pi/T_b} A e^{j\omega kT_b} d\omega = \frac{A}{T_b} \text{Sa}(k\pi) = \begin{cases} A/T_b (\text{常数}) & k = 0 \\ 0 & k \neq 0 \end{cases} \tag{4-6}$$
$$H_{eq}(\omega) = \sum_{i=-\infty}^{\infty} H(\omega + \frac{2i\pi}{T_b}) = C(\text{或常量}) \quad |(\omega)| \leq \frac{\pi}{T_b}$$

结论：

（1）由于理想低通传输滤波器的  $h(t)$  冲激响应为  $\text{Sa}(t)$  函数，也就是说只要系统传输特性  $H(\omega)$  能等效为传输函数  $H_{eq}(\omega)$ ，即在  $|\omega| \leq \pi/T_b$  区间为一等效理想低通滤波器，即可满足无码间串扰的传输条件，消除码间串扰。

从图 4-10 可以看出：当  $t=0$  时， $h(0)=1$ （或常数）；当  $t=kT_b$  ( $k \neq 0$  的整数) 时， $h(kT_b)=0$ 。也就是说，在  $t=T_b、2T_b \cdots$  时刻， $h(kT_b)$  正好为 0，即任一个码元对其后面码元判决时刻不产生干扰。

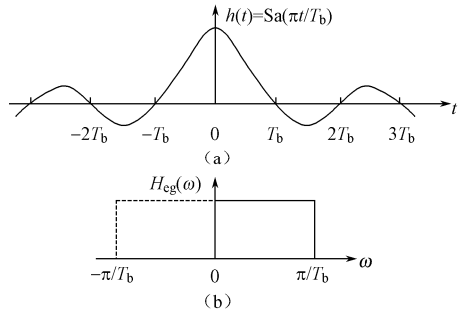


图 4-10 理想低通传输滤波器的冲激响应  $h(t)$

（2）若  $h(t)$  尾巴拖得太长，衰减的较慢，在同步不准时，任一个码元都要对后面几个码元产生串扰，即  $t=kT_b$  ( $k \neq 0$  的整数) 时， $h(kT_b) \neq 0$ ，或者说后面任一个码元都要受到前面几个码元的串扰。因此还要求  $h(t)$  衰减的快一些，尾巴不要拖得太长。

以上两点即为数字基带传输系统无码间串扰的条件。

4.3 基带传输系统的性能改进技术介绍

4.3.1 部分响应技术

1. 引言

在前面的讨论中，为消除码间串扰要求基带系统的总特性  $H(\omega)$  等效为理想低通特性。其响应波形的特点是频谱窄，频带利用率能达到理论上的极限 ( $2B/\text{Hz}$ )，但其缺点是第一个零点以后的尾巴振荡幅度大、收敛慢，从而对定时系统要求十分严格，若定时稍有偏差，则极易引起严重的码间串扰。

那么，能否找到频率利用率既高又使“尾巴”衰减大、收敛快的传输波形呢？下面将证明这种波形是存在的。通常把这种波形称为部分响应波形。利用这种波形进行传送的基带传输系统称为部分响应系统。

部分响应系统最主要的特点在频带利用率达到理论上的极限条件下，可以获得快速衰减大、收敛快的系统响应波形。

为说明部分响应波形的一般特性，让我们先从一个实例谈起，我们已经熟知，理想矩形传输特性的冲激响应波形为

$$\text{Sa}(2\pi\omega t) = \frac{\sin 2\pi\omega t}{2\pi\omega t}$$

现在我们让两个时间上相隔一个码元时间  $T_b$  的波形相加，

$$\frac{\sin 2\pi\omega\left(t + \frac{T_b}{2}\right)}{2\pi\omega\left(t + \frac{T_b}{2}\right)} \text{ 与 } \frac{\sin 2\pi\omega\left(t - \frac{T_b}{2}\right)}{2\pi\omega\left(t - \frac{T_b}{2}\right)}$$

波形相加，如图 4-11 (a) 所示，则相加后的波形为  $g(t)$ 。

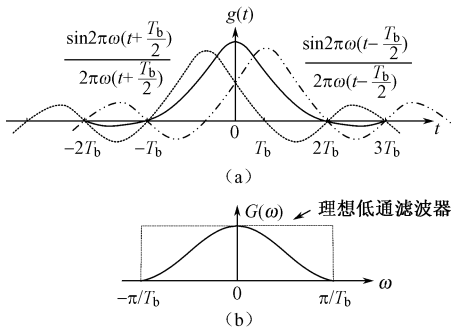


图 4-11 波形  $g(t)$  及频谱  $G(\omega)$

$$g(t) = \frac{\sin 2\pi\omega\left(t + \frac{T_b}{2}\right)}{2\pi\omega\left(t + \frac{T_b}{2}\right)} + \frac{\sin 2\pi\omega\left(t - \frac{T_b}{2}\right)}{2\pi\omega\left(t - \frac{T_b}{2}\right)} = \frac{4}{\pi} \left[ \frac{\cos \pi t / T_b}{1 - 4t^2 / T_b^2} \right] \tag{4-7}$$

可见

$$\begin{cases} g(0) = \frac{4}{\pi} \\ g\left(\pm \frac{T_b}{2}\right) = 1 \\ g\left(\frac{kT_b}{2}\right) = 0, k = \pm 3, \pm 5 \dots \end{cases} \quad (4-8)$$

不难求得  $g(t)$  的频谱函数  $G(\omega)$

$$G(\omega) = \begin{cases} 2T_b \cos \frac{\omega T_b}{2} & |\omega| \leq \pi/T_b \\ 0 & |\omega| > \pi/T_b \end{cases} \quad (4-9)$$

### 2. 部分响应波形的特点

(1) 合成后  $g(t)$  的波形的振荡衰减加快了，原因是相距一个码元间隔的  $\sin x/x$  振荡正负相反而相互抵消。

(2) 根据  $g(t)$  计算公式，可以看出合成波形幅度随  $t$  按  $1/t^2$  变化，即尾巴幅度与  $t^2$  成反比，这说明它比  $\sin x/x$  波形尾部收敛快、衰减也大。

(3) 频带宽度保持不变，而频谱函数变为滚降波形。

(4) 若用  $g(t)$  作为传送波形，且传送码元间隔为  $T_b$ ，则在本码元抽样时刻上仅有前一码元的串扰，而与其他码元不发生串扰（见图 4-12），表面看来，由于前后码元的串扰很大，故似乎无法按  $1/T_b$  的速率进行传送，但仔细分析，我们会发现这时的“串扰”仅仅由前一码元的值确定，而在本码元判断时，前一码元已确定，故实际上可以容易的去掉该“串扰”，所以，在不增加系统带宽条件下，仍然可以按照极限码元速率  $1/T_b$  传送。

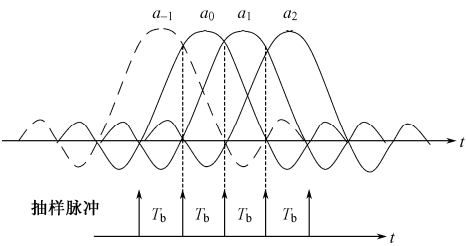


图 4-12 码元串扰示意图

(5) 部分响应技术是人为地在一个以上的码元区间引入一定数量的码间串扰，或者说，在一个以上码元区间引入一定相关性，进行相关编码。

因为这种串扰是人为的、有规律的，所以在接收端可以去掉，不会造成真正的串扰。为了消除差错传播现象，首先进行预编码，而后再进行部分响应编码（也叫相关编码）。

### 3. 部分响应的一般时域表示式

$$\begin{aligned} g(t) = & n_0 \cdot \text{Sa}\left(\frac{\pi}{T_b} t\right) + n_1 \cdot \text{Sa}\left[\frac{\pi}{T_b}(t - T_b)\right] + \\ & \dots + n_{N-1} \cdot \text{Sa}\left[\frac{\pi}{T_b}[t - (N-1)T_b]\right] \end{aligned} \quad (4-10)$$

- (1) 当  $n_0=1$ , 其余为 0 时, 为理想低通型的第 0 类部分响应函数;
- (2) 当  $n_0=0, n_1=1, n_2=1$ , 其余为 0 时, 为第 I 类部分响应函数;
- (3) 当  $n_0=1, n_1=2, n_2=1$ , 其余为 0 时, 为第 II 类部分响应函数;
- (4) 当  $n_0=2, n_1=1, n_2=-1$ , 其余为 0 时, 为第 III 类部分响应函数;
- (5) 当  $n_0=1, n_1=0, n_2=-1$ , 其余为 0 时, 为第 IV 类部分响应函数。
- (6) 当  $n_0=-1, n_1=0, n_2=2, n_3=0, n_4=-1$ , 其余为 0 时, 为第 V 类部分响应函数。

常见的上述部分响应波形定义、波形、频谱示于表 4-1。从表看出, 前面讨论的例子属于 I 类, 各类部分响应波形的频谱均不超过理想低通信号的频谱宽度, 各类  $g(t)$  的频谱在  $1/2T_b$  处为零, 但它们的频谱结构和对邻近码元抽样时刻的串扰不同。并且有  $G(\omega)$  在零频率处也出现零点 (见 IV、V 类)。

表 4-1 常见的部分响应系统

| 类别  | $n_0$ | $n_1$ | $n_2$ | $n_3$ | $n_4$ | $g(t)$ | $ G(\omega) ,  \omega  \leq \pi/T_b$ | 二进制输入时抽样的电平数 |
|-----|-------|-------|-------|-------|-------|--------|--------------------------------------|--------------|
| 0   | 1     |       |       |       |       |        |                                      | 2            |
| I   | 1     | 1     |       |       |       |        |                                      | 3            |
| II  | 1     | 2     | 1     |       |       |        |                                      | 5            |
| III | 2     | 1     | -1    |       |       |        |                                      | 5            |
| IV  | 1     | 0     | -1    |       |       |        |                                      | 3            |
| V   | -1    | 0     | 2     | 0     | -1    |        |                                      | 5            |

目前应用最多的是第 I 类和第 IV 类。第 I 类频谱主要集中在低频段, 适于信道频带高频严重受限的场合。第 IV 类无直流成分, 且低频分量很小。由表 4-1 还可以看出, 第 I、IV 类的抽样电平数 (加权系数  $n_N$  的取值) 比其他几类均少, 这也是它们得到广泛应用的原因之一。

### 4.3.2 时域均衡技术

#### 1. 为什么要均衡

基带系统传输的信号, 一般由一些不同频率的分量组成。若对信号不同的频率分量, 在系统输入输出前后, 各分量间仍保持原有比例的关系, 则信号波形经传输后不产生畸变。但

实际系统和信道的非理想、衰耗随频率的变化而变化等因素，一定会使信号产生畸变，带来码间串扰。

在基带传输中，码间串扰是影响传输质量的主要因素之一。尽管在设计系统时，我们严格按照基带传输系统无码间串扰的要求设计系统，但实际传输特性总是会有偏离或差异，要进一步克服这种实际特性的偏离或差异，就要采用均衡的办法，改善实际系统特性。

2. 什么叫均衡

均衡是指对传输系统的幅频特性和相频特性进行的校正。

采用均衡技术调整接收信号各频率成分的幅度和相位（调整相位相当于调整时延），就能使信道具有理想的传输特性，具体措施是在设备中加均衡器。

均衡器的设计与信号性质有关，对传输电话信号，由于人耳对相位不敏感，所以只考虑传输信道的幅频特性均衡就够了；传输电视信号时，对传输信道的幅频、相频特性均有要均衡，否则图像就失真。对数字信号的传输，也对幅频、相频特性都有要均衡。

均衡技术可分为频域均衡（包括幅度均衡、相位或时延均衡）和时域均衡。前者是校正频率特性，后者则直接校正畸变波形。按调节的方法还可分为固定均衡和可变均衡。可变均衡又可分为手动均衡和自适应（自动）均衡。目前数字信号基带传输系统中主要采用时域均衡，所以我们简单介绍一下时域均衡的原理与技术。

3. 时域均衡的原理与技术

设插入时域均衡滤波器的频率特性为  $T(\omega)$ ， $T(\omega)$ 插入位置如图 4-13 所示，则当这个包括  $T(\omega)$ 在内的系统要消除码间串扰时，必须满足

$$H_{\text{eg}}(\omega) = \sum_{i=-\infty}^{\infty} H\left(\omega + \frac{2\pi}{T_b}i\right) \cdot T\left(\omega + \frac{2\pi}{T_b}i\right) = C \text{ (常数)}$$

(4-11)

$$\omega \text{ 的积分区间为 } |\omega| \leq \frac{\pi}{T_b}$$

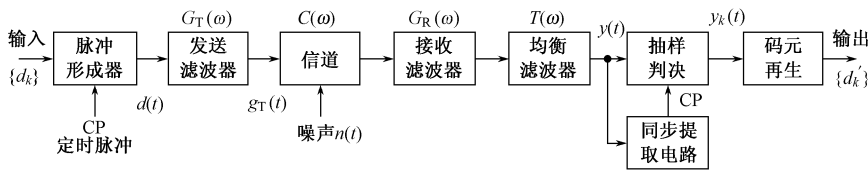


图 4-13 时域均衡器  $T(\omega)$ 插入位置

如果  $\Pi[\omega+(2\pi i/T_b)]$ 对不同的  $i$  有相同的函数形式，即  $T(\omega)$ 为以周期  $2\pi/T_b$  的周期函数，则当  $T(\omega)$ 在  $(-\pi/T_b, \pi/T_b)$  内有

$$T(\omega) = \frac{C}{\sum_i H\left(\omega + \frac{2\pi i}{T_b}\right)}$$

$$|\omega| \leq \frac{\pi}{T_b}$$

(4-12)

既然  $T(\omega)$ 是可以表示为周期为  $2\pi/T_b$  的函数，则  $T(\omega)$ 可用傅里叶级数表示，即

$$T(\omega) = \sum_{n=-\infty}^{\infty} C_n e^{jnT\omega_b} = \left[ \sum_{n=-\infty}^{\infty} \frac{T_b}{2\pi} \int_{-\frac{\pi}{T_b}}^{\frac{\pi}{T_b}} T(\omega) e^{-jn\omega T_b} d\omega \right] e^{jnT\omega_b} \quad (4-13)$$

$$C_n = \frac{T_b}{2\pi} \int_{-\frac{\pi}{T_b}}^{\frac{\pi}{T_b}} \frac{C}{\sum_i H\left(\omega + \frac{2\pi i}{T_b}\right)} e^{-jn\omega T_b} d\omega$$

再对  $T(\omega)$  进行傅氏反变换, 则可求出其单位冲激响应  $h_T(t)$  为

$$h_T(t) = F^{-1}[T(\omega)] = \sum_{n=-\infty}^{\infty} C_n \delta(t - nT_b) \quad (4-14)$$

根据  $h_T(t)$  是许多单位冲激响应之和, 可知  $T(\omega)$  网络是由无限多个横向排列的延迟单元和抽头系统数组成, 因此也称为横向滤波器。

它的功能是利用它产生的无限多响应波形之和, 将抽样时刻上有码间串扰的响应波形转换成抽样时刻上无码间串扰的响应波形。由于横向滤波器的均衡原理是建立在响应波形上的, 故把这故把这种均衡称为时域均衡。

由于横向滤波器特性完全取决于各抽头系数  $c_n$ ,  $c_n$  由  $H(\omega)$  决定, 不同的  $c_n$  值将得出不同的  $h_T(t)$  或  $T(\omega)$ , 因此只要各抽头系数是可调的, 则横向滤波器是通用。另外, 如果抽头系数设计成可调的, 也为随时修改系统的时间响应提供了可能条件。

理论上, 无限长的均衡滤波器才能把失真波形完全校正。但因为实际信道中, 一个码元脉冲波形仅对邻近的少数几个码元产生串扰, 故实际上只要有一二十个抽头的滤波器可以了。抽头数太多会给制造和使用带来困难。

实际应用时, 是用示波器观察均衡滤波的输出信号的眼图, 通过反复调整各个增益放大器的  $C_n$ , 使眼圈的“眼”张开最大为止。

#### 4. 有限长横向滤波器的时域均衡

设横向滤波器共有  $2N$  节延迟线, 每节的延迟时间等于码元宽度  $T_b$ , 在各延迟线之间引出抽头共  $2N+1$  个, 如图 4-14 所示。

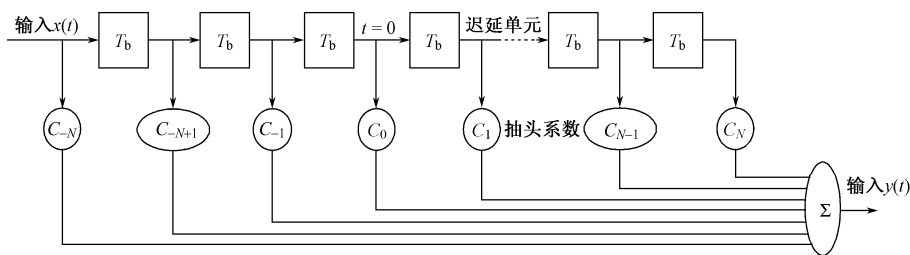


图 4-14 无限多个横向排列的延迟单元组成的横向滤波器

每个抽头的增益可正可负, 因此当输入为有失真的波形  $x(t)$  时, 就可以使相加器输出的信号  $y(t)$  在其他码元抽样时刻的波形的串扰为零, 即

$$y_n = \begin{cases} 1 & n = 0 \\ 0 & n = \pm 1; \pm 2; \pm 3; \dots \pm N \end{cases} \quad \begin{matrix} \text{本码元} \\ \text{其他码元} \end{matrix} \quad (4-15)$$

$$\begin{aligned}
 y_n &= c_{-N} \cdot x_{n+N} + c_{-N+1} \cdot x_{n+N-1} + \dots + c_0 \cdot x_n + \\
 &\quad \dots + c_{N-1} \cdot x_{n-N+1} + c_N \cdot x_{n-N} \\
 &= \begin{cases} 1 & n=0 \\ 0 & n=\pm 1, \pm 2, \dots, \pm N \end{cases}
 \end{aligned} \tag{4-16}$$

用  $n=0 \sim \pm N$  展开上式为  $(2N+1)$  次方程组:

$$\begin{cases} y_{-N} = c_{-N} \cdot x_0 + c_{-N+1} \cdot x_{-1} + \dots + c_0 \cdot x_{-N} + \dots + c_{N-1} \cdot x_{-2N+1} + c_N \cdot x_{-2N} = 0 \\ y_{-N+1} = c_{-N} \cdot x_1 + c_{-N+1} \cdot x_0 + \dots + c_0 \cdot x_{-N+1} + \dots + c_{N-1} \cdot x_{-2N+2} + c_N \cdot x_{-2N+1} = 0 \\ \vdots \\ y_0 = c_{-N} \cdot x_N + c_{-N+1} \cdot x_{N-1} + \dots + c_0 \cdot x_0 + \dots + c_{N-1} \cdot x_{-N+1} + c_N \cdot x_{-N} = 1 \\ \vdots \\ y_{N-1} = c_{-N} \cdot x_{2N-1} + c_{-N+1} \cdot x_{2N-2} + \dots + c_0 \cdot x_{N-1} + \dots + c_{N-1} \cdot x_0 + c_N \cdot x_{-1} = 0 \\ y_N = c_{-N} \cdot x_{2N} + c_{-N+1} \cdot x_{2N-1} + \dots + c_0 \cdot x_N + \dots + c_{N-1} \cdot x_1 + c_N \cdot x_0 = 0 \end{cases}$$

写成矩阵形式:

$$\begin{bmatrix} x_0 & x_{-1} & \dots & x_{-2N} \\ x_1 & x_0 & \dots & x_{-2N+1} \\ x_2 & x_1 & \dots & x_{-2N+2} \\ \vdots & \vdots & & \vdots \\ x_{2N} & x_{2N-1} & \dots & x_0 \end{bmatrix} \cdot \begin{bmatrix} c_{-N} \\ c_{-N+1} \\ \vdots \\ c_0 \\ \vdots \\ c_{N-1} \\ c_N \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 1 \\ \vdots \\ 0 \\ 0 \end{bmatrix} \tag{4-17}$$

简写成

$$X \cdot C = I$$

解这个  $(2N+1)$  独立的线性方程, 就可求得需要  $(2N+1)$  个  $c_k$ 。

时域均衡的基本思想可用图 4-15 所示波形来简单说明。它是利用波形补偿的方法将失真的波形直接加以校正, 这可以利用观察波形的方法直接调节。

**例 4-1** 设有一个三抽头的时域均衡器, 输入波形如图 4-15 (a) 所示, 其中  $x_{-2}=0.0$ ,  $x_{-1}=0.2$ ,  $x_0=1.0$ ,  $x_1=-0.3$ ,  $x_2=0.1$ , 并当  $|n| > 2$  时,  $x_k=0$ 。求其抽头加权系数  $c_k$ 。

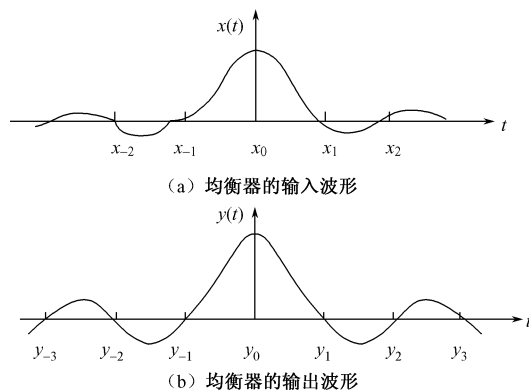


图 4-15 时域均衡器输入与输出波形

**解** 因三抽头则  $2N+1=3$ , 有三个独立的线性方程。



$N=1$ ，抽头系数为  $c_{-1}$ ， $c_0$ ， $c_1$ 。根据已知输入信号条件，有方程如下

$$x_0 c_{-1} + x_{-1} c_0 + x_{-2} c_1 = 0$$

$$x_1 c_{-1} + x_0 c_0 + x_{-1} c_1 = 1$$

$$x_2 c_{-1} + x_1 c_0 + x_0 c_1 = 0$$

代入已知信号  $x$  的值：

$$1.0 \cdot c_{-1} + 0.2 \cdot c_0 + 0.0 \cdot c_1 = 0$$

$$-0.3 \cdot c_{-1} + 1.0 \cdot c_0 + 0.2 \cdot c_1 = 1$$

$$0.1 \cdot c_{-1} - 0.3 \cdot c_0 + 1.0 \cdot c_1 = 0$$

解方程组得：

$$c_{-1} = -0.1779$$

$$c_0 = 0.8897$$

$$c_1 = 0.2847$$

再将均衡器三个抽头系数  $c_k$  的值代入输出  $y_n$  波形的计算公式，核对均衡器的输出。

$$y_n = c_{-N} \cdot x_{n+N} + c_{-N+1} \cdot x_{n+N-1} + \dots + c_0 \cdot x_n + \dots + c_{N-1} \cdot x_{n-N+1} + c_N \cdot x_{n-N}$$

$$= c_{-1} \cdot x_{n+1} + c_0 \cdot x_n + c_1 \cdot x_{n-1}$$

$$= -0.1779 \cdot x_{n+1} + 0.8897 \cdot x_n + 0.2847 \cdot x_{n-1}$$

$$y_{-3} = 0.0$$

$$y_{-2} = -0.0356$$

$$y_{-1} = 0.0$$

$$y_0 = 1$$

$$y_1 = 0.0$$

$$y_2 = -0.0153$$

$$y_3 = 0.0284$$

则三抽头的时域均衡器，输出波形如图 4-15 (b) 所示。

结论：通过三抽头的横向滤波器可以在  $y_{-1} \sim y_1$  范围内消除码间串扰，在这个范围外虽然仍可能有码间串扰，但串扰值很小。只要适当增加横向滤波器的长度就可以将串扰值减到最小。

## 5. 维特比自适应时域均衡器

自适应均衡器所追求的目标就是要达到最佳抽头增益系数，能直接从传输的实际数字信号中获取信息，根据某种均衡算法不断调整抽头系数，以适应信道的随机变化，使均衡器总是保持最佳的工作状态，有较好的失真补偿性能。自适应均衡器有以下三个特点：

- (1) 快速初始收敛特性；
- (2) 好的跟踪信道时变特性；
- (3) 低的运算量。

按最小均方误差准则来构成的维特比 (Viterbi) 均衡器就具有上述特点，限于篇幅，下面仅介绍其总思路。

首先，自适应均衡器在正式工作前先发一定长度的测试脉冲序列，又称训练序列，以调整均衡器的抽头系数，使均衡器基本上趋于收敛，然后再自动改变为自适应工作方式，使均衡器维持最佳状态。

数据信号序列是以突发脉冲串的形式传输的，在突发脉冲串的中部，若加有已知方式的且自相关性强的训练序列，利用这一训练序列，建立起该信道模型的均衡器。这个均衡器随信道特性与时间改变，但在一个突发脉冲串期间被认为是恒定的。

建立了该信道模型的均衡器后，将全部序列再馈入刚建立信道（含最新均衡器），则将含有与训练序列最接近的输出序列，被认为是当前发送的序列。但是当序列长度较大时，产生全部可能的序列将导致很大的时延。

## 4.4 眼图

### 1. 什么是眼图

评价基带传输系统性能的一种定性而方便的方法是观察接收端的基带信号波形。如果将接收波形输入示波器，把产生水平扫描的锯齿波周期与码元定时同步，则在示波器屏幕上可以观察到类似人眼的图案，称之为“眼图”。

### 2. 观察眼图时示波器的连接方法

观察眼图时示波器的连接方法如图 4-16 所示。

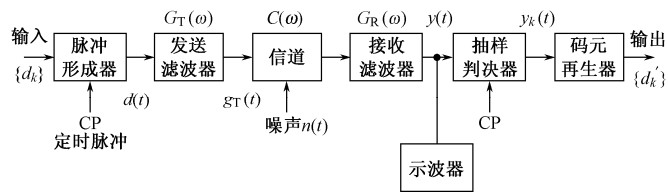


图 4-16 观察眼图时示波器的连接方法

### 3. 无噪声无码间串扰眼图

当数字基带系统为理想低通型，系统不存在噪声，其示波器输入信号为图 4-17（a）二元码时，调整示波器的扫描周期与码元周期  $T_b$  相等，示波器屏幕上可以观察到“眼图”，如图 4-17（b）所示。

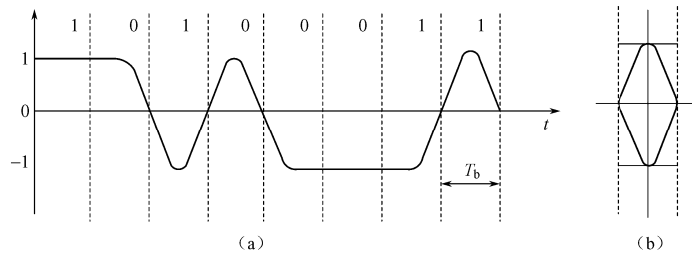


图 4-17 无噪声无码间串扰时眼图

### 4. 无噪声有码间串扰时波形及眼图

当数字基带系统为非理想低通型，系统不存在噪声，其示波器输入信号为图 4-18（a）二元码时，调整示波器的扫描周期与码元周期  $T_b$  相等，示波器屏幕上可以观察到“眼图”，

如图 4-18 (b) 所示。

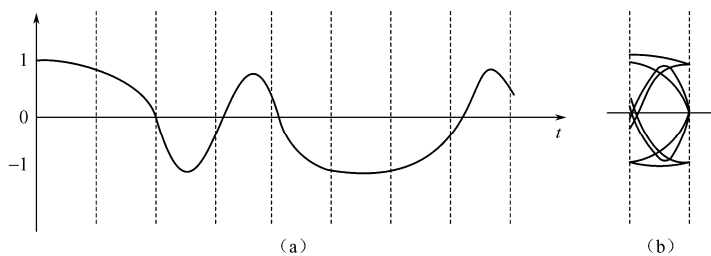


图 4-18 无噪声有码间串扰时眼图

5. 有噪声有码间串扰时眼图

当数字基带系统为非理想低通型，系统也存在噪声，其示波器输入信号为图 4-19 (a) 二元码时，调整示波器的扫描周期为码元周期  $T_b/4$ ，示波器屏幕上可以观察到“眼图”，如图 4-19 (b) 所示。

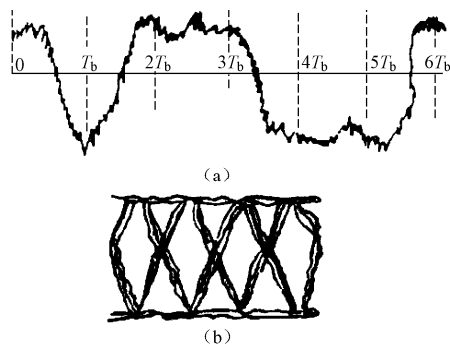


图 4-19 有噪声有码间串扰时眼图

眼图对数字基带信号传输系统的性能给出了很多有用的信息，可以从中看出码间串扰的大小和噪声的大小。其中：

- (1) 噪声越大，眼图边沿带状线越模糊；
- (2) 码间串扰越大，眼图越不端正。

6. 眼图的一般模型

可以把眼图和数字基带系统的性能关系，简化为图 4-20 所示的形状，称其为眼图的一般模型。从图 4-20 中可以得出：

- (1) 最佳抽样时刻应选择在眼图中眼睛张开的最大处；
- (2) 对定时误差的灵敏度，由斜边斜率决定，斜率越大，对定时误差就越灵敏；
- (3) 在抽样时刻上，眼图上、下两分支的垂直高（宽）度，都表示最大信号畸变；
- (4) 在抽样时刻上，眼图上、下两分支，离门限电平最近的一根线迹，至门限的距离表示各相应电平的噪声容限，噪声瞬时值超过它就可能发生判决差错；
- (5) 对于从信号过零点取平均来得到定时信息的接收系统，眼图倾斜分支与横轴相交的区域的大小，表示零点位置的变动范围，这个变动范围的大小对提取定时信息有重要的影响。

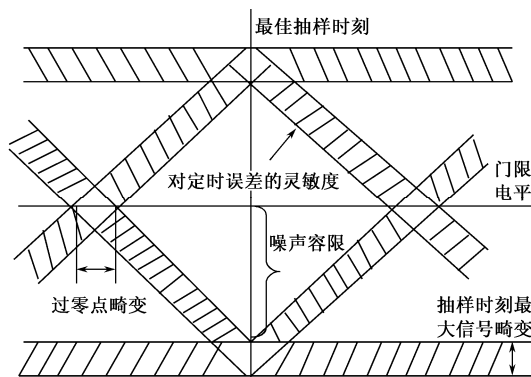


图 4-20 眼图的一般模型

总之，通过示波器直观系统的眼图，可以帮助调整系统中的发送滤波器、接收滤波器，以减小码间串扰。

## 思考与练习

- 4-1 什么是基带信号？基带信号有哪几种常用的形式？
- 4-2 设二进制符号序列为 110010001110，试以矩形脉冲为例，分别画出相应的单极性码、双极性码、单极性归零码、双极性归零码、二进制差分码。
- 4-3 已知信息代码为 10000000011，求相应的 AMI 码和 HDB3 码。
- 4-4 什么叫码间串扰？它是怎样产生的？有什么不好的影响？应该怎样消除或减小？
- 4-5 能满足无码间串扰条件的传输特性冲激响应  $h(t)$  是怎样的？为什么说能满足无码间串扰条件的  $h(t)$  不是唯一的？
- 4-6 什么叫眼图？它有什么用处？为什么双极性码与 AMI 码的眼图具有不同形状？
- 4-7 设随机二进制脉冲序列的码元间隔为  $T$ ，经过理想抽样以后，送到图 4-21 所示的几种滤波器，指出哪几种会引起码间串扰，哪几种不会引起码间串扰。

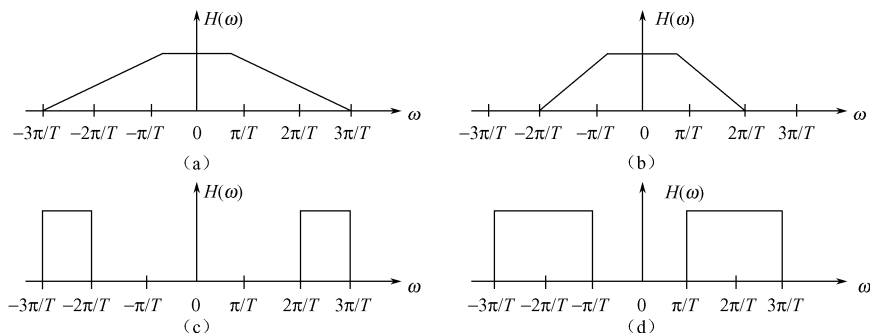


图 4-21

- 4-8 传送两种码元速率  $R_B=1\,000$  和  $1\,500$  (Baud) 的数字基带信号，试问系统采用图 4-22 所画的哪一种传输特性较好？并简要说明其理由。

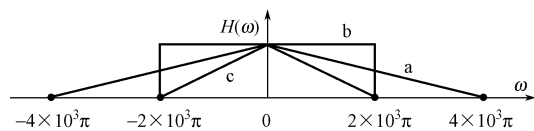


图 4-22

4-9 试画出 1110010011010 的眼图 (码元速率为  $f_b=1/T_b$ )。

(1) “1” 码用  $g(t)=[1+\cos(\pi t/T_b)]/2$  表示, “0” 码用  $-g(t)$  表示;

(2) “1” 码用  $g(t)=[1+\cos(\pi t/T_b)]/2$  表示, “0” 码用 0 表示。

4-10 部分响应系统实现频带利用率为  $2B/\text{Hz}$  的原理是什么?

4-11 时域均衡怎样改善系统的码间串扰?

## 第5章 高性能数据调制技术

信号调制即对信号的调和整治,使信号匹配信道特性以适合信道传输。调制可实现频率分配、多路复用、减少噪声干扰以及方便辐射等。在信号调制系统中,通常把待传输消息对应的基带信号称为调制信号,调制系统输出的信号称为已调制信号,简称已调信号,用于承载基带信号传输的光波或电波称为载波。解调是调制的逆过程,就是从已调信号中提取消息信号的过程,这样可方便接收者(信宿)处理或解释。本章主要讨论最小频移键控(MSK)、正交振幅调制(QAM)、正交频分复用(OFDM)等高性能调制技术。

### 5.1 引言

#### 5.1.1 调制技术的主要性能指标

##### 1. 数据调制技术的主要性能指标

调制技术的性能指标有多种,因通信系统的实际要求的不同会有所不同,例模拟信号调制解调增益(制度增益),是评价其调制解调系统性能好坏的重要指标,但数据调制技术一般不用,通常采用功率利用率和带宽利用率。

功率利用率

$$\eta_p = E_b / n_0 \quad (5-1)$$

其中,  $E_b$  为每比特信息码的平均能量,  $n_0$  为白噪声单边功率谱密度。

在比特差错率相同(例如  $10^{-6}$ )的不同调制系统中,归一化信噪比  $E_b/n_0$  越小,说明此系统的功率利用率越好。

功率利用率描述了在低功率情况下,一种调制技术保持数字信息正确传输的能力。

频带利用率也叫带宽利用率,它描述了在有限的带宽内调制系统的数据传输能力。

一般而言,增加数据率就意味着减小数字码元的脉冲宽度,即增加信号的带宽。因此,数据率和带宽有之间存在着一种矛盾关系。

由此可见,带宽利用率给出的是如何有效利用所分配带宽的一种测度,其定义为:单位频带(1 Hz)内所能实现的信息速率(或码元速率)。用  $R_b/B$  (或  $R_B/B$ ) 表示:

$$\eta_b = R_b/B \quad (5-2)$$

其中  $R_b$  表示传信率,  $R_B$  表示传码率,  $B$  表示系统带宽。

如果一种调制方式的  $R_b/B$  值大,那么说明在分配的带宽内传输的数据多,频带利用率就高。

根据香农(Shannon)信道容量定理可知,数字调制信道的码元极限速率为  $C = B \log_2(1+S/N)$ ,理论上,只要信息比特速率  $R_b \leq C$ ,通过采用足够复杂的纠错编码方式,该信道就能以任意小的差错概率,进行速率为  $R_b$  的信息传输;若  $R_b > C$ ,则不可能存在某种纠错编码方式使传输差错概率任意小。

对于任意小的错误概率，最大可能的带宽利用率为

$$\eta_{b\max} = \frac{C}{B} = \log_2 \left( 1 + \frac{S}{N} \right) = \log_2 \left( 1 + \frac{S}{n_0 \cdot B} \right) \quad (5-3)$$

结论：最大可能的频带利用率与信号功率成正比、与信道噪声和信号带宽成反比。

功率利用率与频带利用率的关系：在差错控制编码中，通过增加冗余度提高了差错控制能力，使信号接收功率可以降低，但增加冗余度同时使信息比特带宽增加，即降低了频带利用率。在数据通信系统设计中，经常需要在两个指标之间折中。

功率利用率和频带利用率是调制系统中较为重要的指标，但不同的实际系统有自己认为重要的性能指标。如包络特性、适应信道变化特性和实现复杂度等。

## 2. 无线信道的特征及对数据调制技术基本要求

无线通信信道的基本特征是：

- (1) 因受频率资源和信道的传播特性限制，无线信道带宽有限；
- (2) 无线通信工作在完全暴露的电磁环境下，因此，所受到干扰和噪声的影响大；
- (3) 无线移动信道是随参信道，存在多径衰落。

无线数据调制技术基本要求如下：

- (1) 无线数据通信系统的设计者在选择调制方式时，必须考虑信道带宽有限所造成影响和限制。换句话说，确定数据调制信号的功率谱密度就显得十分重要。
- (2) 针对无线通信信道的特点，已调信号应具有高的频谱利用率、抗噪声和抗干扰能力强、易于采用相干或非相干方式解调、以及适合在衰落信道中传输等特点。
- (3) 高的频谱利用率要求已调信号所占的带宽窄，即已调信号频谱主瓣要窄，同时副瓣的幅度要低（即辐射到相邻频道的功率要小）。
- (4) 高的抗干扰和抗多径衰落性能，要求在恶劣的移动信道环境下能很好地工作，经过调制解调后的输出信噪比（ $S/N$ ）较大或误码率较低。

### 5.1.2 无线调制技术的发展与分类

在介绍调制技术的发展前，先说明相位路径的概念。

由于  $\omega = d\theta(t)/dt$ ，因此一个已调波的频谱特性与其相位路径有着密切的关系。所谓相位路径，是指载波相位  $\theta(t)$  随时间变化的轨迹。不同的调制方式，具有不同的相位路径。

采用 FSK 或 PSK 传递信息时，由于频率和相位之间互为微分与积分的关系，从本质上来说，FSK 也可以看成是 PSK。因此，这类调制方式都可以认为是由载波相位的变化值来传递信息。即随着输入数据的不断变化，载波相位的变化值也在不断变化。换句话说，载波相位的变化值是一个随时间而变化的函数。

回顾调制技术的历史发展过程，为了控制已调波的频谱特性以获得较好的性能，就必须控制它的相位路径。20 世纪 50 年代末出现了二相相移键控（BPSK）调制，后来为了提高信道频带利用率，在它的基础上提出了四相相移键控（QPSK）调制。它们产生的已调波在码元转换时刻都可能产生  $180^\circ$  的相位突跳，使得频谱中高频滚降缓慢，带外辐射大。为消除  $180^\circ$  的相位突跳，改善频谱特性，到 20 世纪 60 年代末又在 QPSK 的基础上提出了交错四相相移键控（OQPSK）。它虽然克服了  $180^\circ$  相位突跳的问题，但是在码元转换时刻仍可能有  $90^\circ$  的相

位突跳，同样使得频谱中高频不能很快的滚降。为了彻底解决相位突跳的问题，人们自然会想到相邻码元之间的相位变化不应该瞬时的突变，而应该在一个码元时间内逐渐累积来完成，从而保持码元在转换时刻相位连续。这就是在 20 世纪 70 年代初提出的最小频移键控(MSK)，其相位累积规律是直线形变化。由于其相位累积仅仅局限于一个码元内进行，其大大限制了选择不同相位路径的可能性；因此，想到把相位路径的研究扩展到几个码元内进行，1977 年提出了受控调频（TFM）。通过采用高斯滤波器来代替 TFM 中的相关编码器，在 1979 年提出了调制前高斯滤波的最小频移键控（GMSK）。为了进行简单的比较，表 5-1 列出了几种在无线通信系统中使用的调制技术的相位路径特点。

表 5-1 几种在无线通信系统中使用的调制技术的相位路径

| 名 称            | 相位路径特点                     | 提 出 时 间     |
|----------------|----------------------------|-------------|
| BPSK           | $0, \pi$ （突跳）              | 20 世纪 50 年代 |
| QPSK           | $0, \pm\pi/2, \pi$ （突跳）    | 20 世纪 60 年代 |
| OQPSK          | $0, \pm\pi/2$ （突跳）         | 20 世纪 70 年代 |
| MSK            | 直线（连续）                     | 20 世纪 70 年代 |
| GMSK           | 非直线（连续）                    | 1979 年      |
| $\pi/4$ -DQPSK | $\pm\pi/4, \pm3\pi/4$ （突跳） | 20 世纪 80 年代 |

目前，在实际无线通信系统中，有 3 类用得最多的数据调制方式：线性调制技术；恒定包络（连续相位）调制技术；幅度、相位和频率的联合调制技术。

1. 线性调制技术

线性调制技术中所谓的“线性”，是指这类调制技术要求通信设备从频率变换、放大直到发射的整个过程中保持充分的线性。显然，这种要求在制造设备时会增加难度和成本，但这类调制技术可获得较高的频谱利用率。此类调制技术主要包括 PSK，QPSK，DQPSK，OQPSK， $\pi/4$ -DQPSK 和多电平 PSK 等。

2. 恒定包络（连续相位）调制技术

什么是恒包络调制？

在模拟调频 FM、调相 PM 和连续相位数据调制 CPFSK 中，已调信号的幅度是不变的，称此为恒包络调制，恒定包络调制的已调信号相位连续，故也称连续相位调制。恒定包络（连续相位）调制技术是泛指那些载波相位以连续形式变化（已调信号包络恒定）的一大类频率调制技术。

为什么要采用恒包络调制技术？

在非 CPFSK 数据调制方式中，当基带码元为矩形时，矩形波的带宽是无限的，而实际电路的带宽总是有限的，并且为了减少码间干扰，通常用带通滤波器限制其带宽，经过这种处理的信号将不再具有恒包络的特性。

另外，在许多无线系统（如在移动无线手机）中使用的高功率放大器，通常工作于非线性状态。而这些非线性器件具有“幅相转换（AM/PM）”效应，即当输入信号幅度变化时，能转换为输出信号的相位变化，造成频谱扩散现象，已调信号出现包络起伏。

恒定包络调制技术的优点是可以使用功率效率高的放大器，而不会使发送信号占用的频



谱增大，即已调信号具有相对窄的功率谱且对放大设备没有线性要求；带外辐射低，可达 $-60\sim-70\text{ dB}$ ；允许采用限幅的方式消除由于干扰引起的幅度变化，简化接收机的设置，并能很好地抵抗随机噪声和由 Rayleigh 衰落引起的信号波动，从而具有较高的抗干扰能力。不足之处是占用带宽比线性调制大，即频谱利用率通常低于线性调制技术。

现代恒定包络数据调制技术主要包括 MSK，GMSK，GFSK 和 TFM 等。提高频谱利用率是提高通信容量的重要措施，是人们规划和设计通信系统的焦点。在 20 世纪 80 年代初期，当时人们选择数字调制技术时，大多把注意力集中于恒定包络数字调制（如泛欧 GSM 蜂窝系统采用的 GMSK）技术上。但是在 20 世纪 80 年代中期以后，人们却着重采用 QPSK 之类的线性数字调制技术（如美国的 IS-54 蜂窝系统和日本的 PDC 蜂窝系统均采用 $\pi/4$ -DQPSK，美国的 IS-95 蜂窝系统在上、下行链路中分别采用 OQPSK 和 QPSK）。

### 3. 幅度、相位与频率联合调制技术

单独使用振幅、相位或载波携带信息时，不能最充分地利用信号平面，若将幅度、相位与频率联合调制，不但可以提高码元传输速率，还可以提高抗干扰和抗衰落的能力。

例如，4 电平、16 电平、64 电平以至 256 电平的幅度和相位联合调制已在微波通信中获得成功应用。以往，人们认为多电平 QAM 信号的特征不适于在移动环境中进行传输。近几年，随着研究工作的深入，人们提出不少改进方案。例如，根据移动信道特征的好坏可自适应地改变 QAM 的电平数，即改变信道传输速率，从而构成变速率 QAM（VR-QAM）调制技术。为减少码间干扰和时延扩展的影响，把将要传输的数据流划分为若干个子数据流，每个子数据流具有低得多的传输速率，并且用这些子数据流去调制若干个载波，从而形成多载波 QAM（MC-QAM）调制技术或正交频分复用（OFDM）调制等，该技术已被 4G 移动通信采用。

## 5.2 最小频移键控（MSK）

### 5.2.1 最小频移键控的基本原理

最小频移键控（Minimum Frequency Shift Keying，MSK）是一种相位连续的二进制调频（CP2FSK）调制的改进型，它是一种典型的恒定包络（连续相位）调制，其频谱利用率高、误比特率低和自同步性能好，适合在无线通信系统中使用。

下面从 MSK 信号、MSK 调制和 MSK 解调方面介绍 MSK 的原理。

#### 1. MSK 信号的数学描述

MSK 是 FSK 的改进型，二进制 MSK 信号的表示式可写为

$$s_{\text{MSK}}(t) = A\cos\left[\omega_c t + \frac{a_k\pi}{2T_b}t + \varphi_k\right] \tag{5-4}$$

或者

$$s_{\text{MSK}}(t) = A\cos[\omega_c t + \theta(t)] \tag{5-5}$$

这里

$$\theta(t) = \frac{\pi a_k}{2T_b} t + \varphi_k \quad (k-1)T_b \leq t \leq kT_b \quad (5-6)$$

式中,  $A$  为载波幅度,  $\omega_c$  为载波频率;  $a_k$  是数据基带信号, 为第  $k$  个码元数据, 取值  $\pm 1$ ;  $T_b$  为比特周期 (这里是码元宽度);  $\theta(t)$  为附加相位,  $\varphi_k$  为第  $k$  个码元起始参考相位, 并在  $(k-1)T_b \leq t \leq kT_b$  时间段内保持不变。

从式 (5-4) 可以推出, 第  $k$  个码元的频率取决于  $a_k$ , 而  $a_k$  的取值为  $\pm 1$ , 故第  $k$  个码元可能出现的两个频率为

$$\begin{aligned} f_1 &= f_c - \frac{1}{4T_b} & \text{当 } a_k = -1 \\ f_2 &= f_c + \frac{1}{4T_b} & \text{当 } a_k = +1 \end{aligned} \quad (5-7)$$

并且

$$f_c = \frac{f_2 - f_1}{2}$$

则频率间隔为

$$\Delta f = f_2 - f_1 = \frac{1}{2T_b} \quad (5-8)$$

调制指数为

$$\Delta f \cdot T_b = \frac{1}{2T_b} \cdot T_b = 0.5 \quad (5-9)$$

对应的 CP2FSK 信号表示式为

$$s(t) = \begin{cases} A \cos(2\pi f_1 t + \varphi_k) & \text{当 } a_k = -1 \\ A \cos(2\pi f_2 t + \varphi_k) & \text{当 } a_k = +1 \end{cases} \quad (5-10)$$

可以证明式 (5-10) 中两个不同频率信号波形具有如下的相关系数

$$\rho = \frac{\sin 2\pi(f_2 - f_1)T_b}{(f_2 - f_1)T_b} + \frac{\sin 4\pi f_c T_b}{4\pi f_c T_b} \quad (5-11)$$

由于 MSK 采用正交调制, 其对应两个不同频率信号波形的相关系数应等于零, 即对应式 (5-11) 右边应恒等于零。

式 (5-11) 右边第一项为零的条件是

$$2\pi(f_2 - f_1)T_b = k\pi \quad k = 1, 2, 3, \dots \quad (5-12)$$

令  $k$  等于最小值 1, 导出的结果如式 (5-8) 所示, 显然式 (5-4) 的 MSK 信号能够使得第一项等于零。

式 (5-11) 右边第二项为零的条件是

$$4\pi f_c T_b = k\pi \quad k = 1, 2, 3, \dots \quad (5-13)$$

即

$$T_b = \frac{k}{4f_c} = \frac{k}{4} T_c \quad k = 1, 2, 3, \dots \quad (5-14)$$

或

$$f_c = \frac{k}{4T_b} = \frac{k}{4} f_b = \left( N + \frac{m}{4} \right) f_b \quad m = 0, 1, 2, \dots \quad (5-15)$$

式 (5-14) 说明, MSK 信号码元周期必须包含有整数倍个四分之一载波周期, 式 (5-15) 说明, MSK 信号载波频率必须包含有整数倍个四分之一码元速率。相应的式 (5-7) 可写成

$$\begin{aligned} f_1 &= f_c - \frac{1}{4T_b} = \left(N + \frac{m+1}{4}\right) \frac{1}{T_b} \\ f_2 &= f_c + \frac{1}{4T_b} = \left(N + \frac{m-1}{4}\right) \frac{1}{T_b} \end{aligned} \quad (5-16)$$

附加相位  $\theta(t)$  的选择应保证信号相位在码元转换时刻连续, 即

$$\theta_{k-1}(kT_b) = \theta_k(kT_b) \quad (5-17)$$

将式 (5-6) 代入式 (5-17) 可以得到

$$\frac{\pi a_{k-1}}{2T_b} t + \varphi_{k-1} = \frac{\pi a_k}{2T_b} t + \varphi_k$$

进一步可以得到

$$\varphi_k = \varphi_{k-1} + (a_{k-1} - a_k) \cdot \frac{k\pi}{2} = \begin{cases} \varphi_{k-1} & \text{当 } a_k = a_{k-1} \\ \varphi_{k-1} \pm k\pi & \text{当 } a_k \neq a_{k-1} \end{cases} \quad (5-18)$$

式 (5-18) 表明: MSK 信号在第  $k$  个码元的起始参考相位  $\varphi_k$ , 不仅于当前的码元  $a_k$  有关, 而且与前面码元  $a_{k-1}$  及起始参考相位  $\varphi_{k-1}$  有关, 也就是说, 先后码元起始参考相位之间存在相关性。

对于相干解调来说,  $\varphi_k$  的起始参考值可以假定为零, 因此式 (5-18) 可以写为

$$\varphi_k = 0 \text{ 或 } \pi \quad (5-19)$$

这是, 进一步分析式 (5-6) 可以看到,  $\theta(t)$  是一个直线函数, 对于第  $k$  个码元, 其持续时间为  $(k-1)T_b \leq t \leq kT_b$ , 则该码元附加相位的初始相位为

$$\theta[(k-1)T_b] = \frac{\pi \cdot a_k}{2T_b} [(k-1)T_b] + \varphi_k = a_k (k-1) \frac{\pi}{2} + \varphi_k \quad (5-20)$$

附加相位的终止相位为

$$\theta(kT_b) = a_k \cdot k \cdot \frac{\pi}{2} + \varphi_k \quad (5-21)$$

因此

$$\Delta\theta_k = \theta(kT_b) - \theta[(k-1)T_b] = a_k \cdot \frac{\pi}{2} = \begin{cases} +\pi/2 & a_k = +1 \\ -\pi/2 & a_k = -1 \end{cases} \quad (5-22)$$

$\theta(t)$  是 MSK 信号的总相位减去随时间线性增长的载波相位  $\omega_c t$  而得到的剩余相位, 称它作附加相位函数。式 (5-22) 表明在每一个码元时间内, 相对前一码元附加相位,  $\theta_k(t)$  不是增加  $\pi/2$ , 就是减小  $\pi/2$ 。

$\theta(t)$  随时间  $t$  的变化规律如图 5-1 所示, 其中正斜率直线表示传 1 码的相位轨迹, 负斜率直线表示传 0 码的相位轨迹, 这种由相位轨迹构成的图称为相位网格图。从图 5-1 中我们可以找出任意一个数据序列由零开始的路径, 例如, 粗线给出了数据序列为 1101000... 的相位路径。

通过以上讨论可知, MSK 信号具有如下特点:

- (1) MSK 信号包络恒定, 即为等幅波;
- (2) 码元转换时, 信号附加相位连续, 即信号的波形没有跳变 (突跳);
- (3) MSK 信号的频偏必须等于  $\pm T_b/4$ , 其调制指数为 0.5;

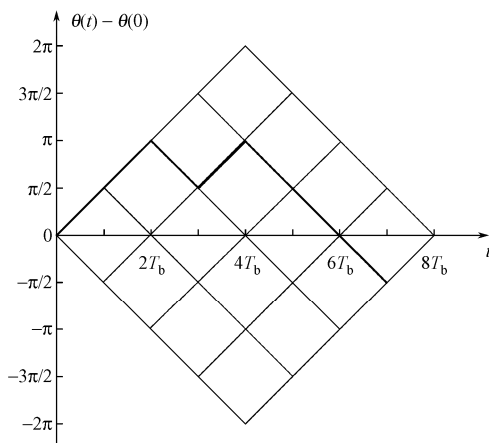


图 5-1 MSK 信号的相位轨迹图

- (4) MSK 信号码元周期必须包含有整数倍个四分之一载波周期；  
 (5) MSK 信号的附加相位  $\theta(t)$  在一个码元内，准确地线性变化  $\pm\pi/2$ 。

## 2. MSK 信号的产生与解调

利用三角公式展开式 (5-4) 得

$$\begin{aligned}
 s_{\text{MSK}}(t) &= A \cos \left[ \omega_c t + \frac{a_k \pi}{2T_b} t + \varphi_k \right] \\
 &= A \left[ \cos \left( \frac{a_k \pi}{2T_b} t \right) \cdot \cos \varphi_k - \sin \left( \frac{a_k \pi}{2T_b} t \right) \cdot \sin \varphi_k \right] \cdot \cos \omega_c t \\
 &\quad - A \left[ \sin \left( \frac{a_k \pi}{2T_b} t \right) \cdot \cos \varphi_k + \cos \left( \frac{a_k \pi}{2T_b} t \right) \cdot \sin \varphi_k \right] \cdot \sin \omega_c t
 \end{aligned} \tag{5-23}$$

假设  $\varphi_k$  起始参考值为零，由式 (5-19) 可知

$$\begin{aligned}
 \cos \varphi_k &= \pm 1 \\
 \sin \varphi_k &= 0
 \end{aligned}$$

则式 (5-23) 可以表示为

$$\begin{aligned}
 s_{\text{MAK}} &= A \left[ \cos \left( \frac{a_k \pi}{2T_b} t \right) \cdot \cos \varphi_k \right] \cdot \cos \omega_c t - A \left[ \sin \left( \frac{a_k \pi}{2T_b} t \right) \cdot \cos \varphi_k \right] \cdot \sin \omega_c t \\
 &= A [I(t) \cdot \cos \omega_c t - Q(t) \cdot \sin \omega_c t]
 \end{aligned} \tag{5-24}$$

在式 (5-24) 中， $I(t)$  为同相分量， $Q(t)$  为正交分量，可以表示为

$$\begin{aligned}
 I(t) &= \cos \left( \frac{a_k \pi}{2T_b} t \right) \cdot \cos \varphi_k = \cos \left( \frac{\pi}{2T_b} t \right) \cdot \cos \varphi_k = a_I(t) \cdot \cos \left( \frac{\pi}{2T_b} t \right) \\
 Q(t) &= \sin \left( \frac{a_k \pi}{2T_b} t \right) \cdot \cos \varphi_k = a_k \cdot \sin \left( \frac{\pi}{2T_b} t \right) \cdot \cos \varphi_k = a_Q(t) \cdot \sin \left( \frac{\pi}{2T_b} t \right)
 \end{aligned} \tag{5-25}$$

式中

$$\begin{aligned} a_k &= \pm 1 \\ a_1(t) &= \cos \varphi_k \\ a_Q(t) &= a_k \cdot \cos \varphi_k \end{aligned}$$

结合式（5-18）分析可以证明， $a_1(t)$ 和  $a_Q(t)$ 每  $2T_b$  输出一对码元，其中  $a_1(t)$ 是数字序列  $a_k$  的差分编码  $c_k$  的奇数位输出， $a_Q(t)$ 是数字序列  $a_k$  的差分编码  $c_k$  的偶数位，并延时  $T_b$  输出。图 5-2 给出了逻辑序列  $d_k=(11010001000111)$  对应的各类波形输出。

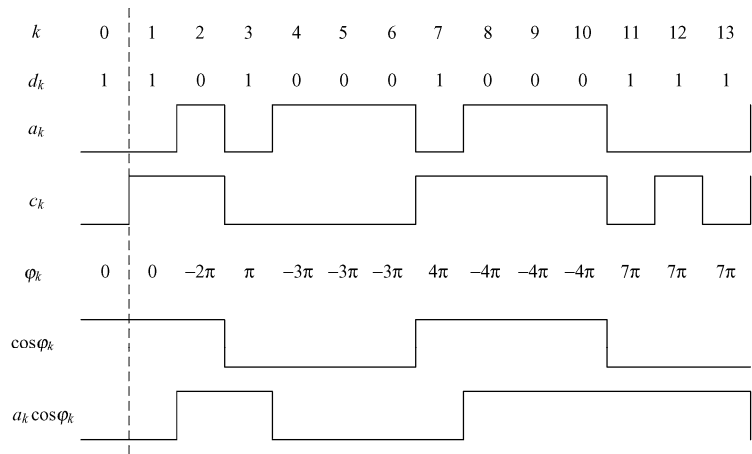


图 5-2 输入数据与各支路数据之间的关系

如果从逻辑上分析图 5-2 可以看出，假设逻辑“1”对应“-”电平，逻辑“0”对应“+”电平，则对于绝对序列  $a_k$  的差分编码  $c_k$  为

$$c_k=(+1+1-1-1-1-1+1+1+1+1-1+1)$$

对应的奇数位输出  $a_1(t)$ 为

$$a_1(t)=(+1-1-1+1+1-1)$$

对应的偶数位输出  $a_Q(t)$ 为

$$a_Q(t)=(+1-1-1+1+1+1)$$

图 5-3 给出了基于式（5-24）和式（5-25）构建的 MSK 信号调制器，与产生过程对应的相干解调器原理如图 5-4 所示。

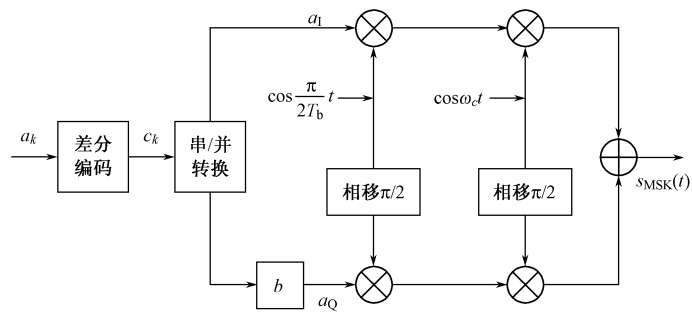


图 5-3 MSK 调制器原理图

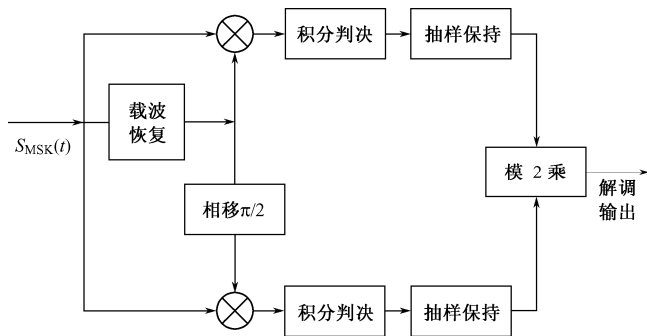


图 5-4 MSK 信号相干解调器原理图

### 3. MSK 信号的功率谱密度

经计算，MSK 信号的功率谱密度如图 5-5 所示。为了比较，也给出了 MSK、GMSK 和 2PSK 调制信号的功率谱密度曲线。

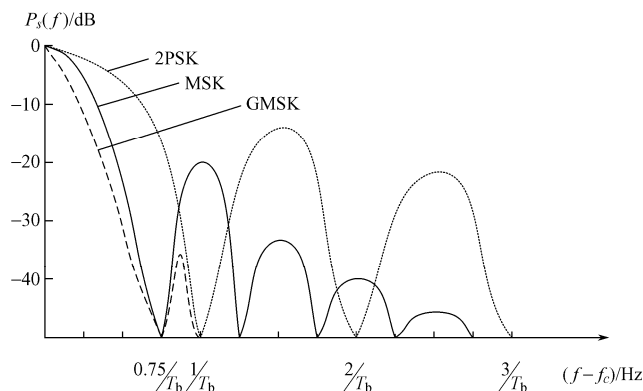


图 5-5 MSK、GMSK 和 2PSK 信号的功率谱密度曲线

设码元周期为  $T_b$ ，则包含 90% 和 99% 的信号功率的带宽的近似值如表 5-2 所示。

表 5-2 部分数字调制信号带宽

| 信号功率百分比 | BPSK    | QPSK    | OQPSK   | MSK       |
|---------|---------|---------|---------|-----------|
| 90 %    | $2/T_b$ | $1/T_b$ | $1/T_b$ | $1/T_b$   |
| 99 %    | $9/T_b$ | $6/T_b$ | $6/T_b$ | $1.2/T_b$ |

#### 5.2.2 高斯滤波最小频移键控（GMSK）

MSK 信号的相位虽然是连续变化的，但在信息代码发生变化时刻，相位变化会出现尖角，即附加相位的导数不连续。这种不连续性降低了 MSK 信号功率谱旁瓣的衰减速度。为了进一步使信号的功率谱密度集中和减小对邻道的干扰，常在 MSK 调制前，对基带信号进行高斯滤波处理，这就是另一种在移动通信中得到广泛应用的恒包络调制方法——调制前高斯滤波的最小频移键控，简称高斯最小频移键控，记为 GMSK，调制方式如图 5-6 所示。

GMSK 的基本原理是让基带信号先经过高斯滤波器滤波，使基带信号形成高斯脉冲，之

后进行 MSK 调制。由于滤波形成的高斯脉冲包络无陡峭边沿，亦无拐点，所以经调制后的已调波相位路径在 MSK 的基础上进一步得到平滑，相位轨迹如图 5-7 所示。

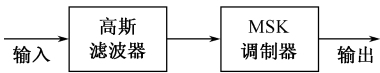


图 5-6 GMSK 调制原理方框图

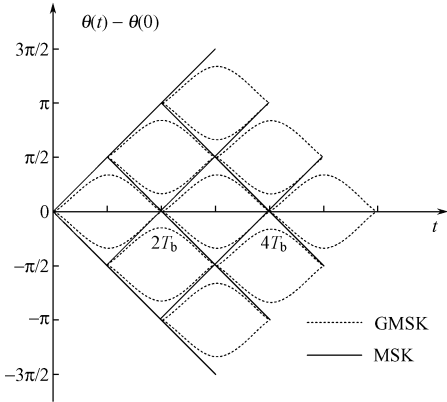


图 5-7 GMSK 相位轨迹图

由图 5-7 可以看出，它把 MSK 信号的相位路径的尖角平滑掉了，因此频谱特性优于 MSK。

### 5.3 正交振幅调制（QAM）

#### 5.3.1 正交振幅调制的基本原理

正交振幅调制（Quadrature Amplitude Modulation，QAM）是用两路独立的基带信号对两个相互正交的同频载波进行抑制载波双边带 ASK 调制，已调制载波的振幅和相位均随两个独立的基带信号变化而变，从而实现两路并行的数字信息的传输。

##### 1. QAM 信号的数学描述

如果设两路的  $M$  进制基带信号  $x(t)$ 和  $y(t)$ 分别为

$$\begin{aligned} x(t) &= \sum_{k=-\infty}^{\infty} x_k \cdot g(t - kT_b) \\ y(t) &= \sum_{k=-\infty}^{\infty} y_k \cdot g(t - kT_b) \end{aligned} \qquad (0 \leq t \leq T_b) \tag{5-26}$$

式中， $T_b$  为多进制码元间隔， $x_k$  和  $y_k$  为双极性  $M$  进制码元，是承载信息的正交载波的信号幅度，例如取为  $\pm 1, \pm 3, \cdots, \pm(M-1)$ 等。则 QAM 信号可表示为

$$s_{\text{QAM}}(t) = x(t) \cdot \cos \omega_c t + y(t) \cdot \sin \omega_c t \tag{5-27}$$

QAM 也可以通过三角函数和差化积，转化为

$$s_{\text{MQAM}}(t) = A_i \cos(\omega_c t + \phi_i) \tag{5-28}$$

式中

$$\begin{aligned} A_i &= \sqrt{x^2(t) + y(t)} \\ \phi_i &= \arctan^{-1}[y(t)/x(t)] \end{aligned}$$

式（5-28）表示 QAM 信号波形是一种振幅和相位联合调制方式。

QAM 信号也可以用星座图表示，图 5-8 给出了二进制 QAM (4QAM)、四进制 QAM (16QAM)、八进制 QAM (64QAM) 对应的星座图。由于其星座图的结构为方形，所以也称为“方形星座图”。

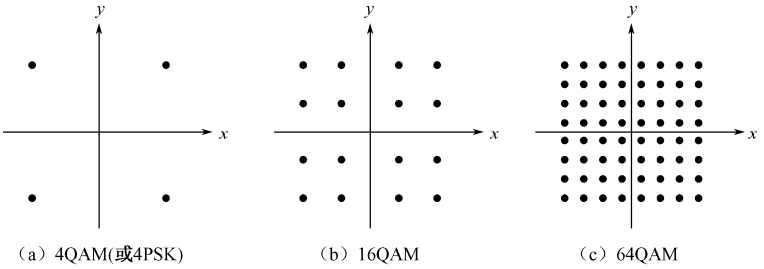


图 5-8 多进制 QAM 方形星座图

对于 4QAM，当 2 路信号幅度相等时，其产生、解调、性能及相位矢量均与 4PSK（或 QPSK）相同。

### 2. QAM 信号调制解调原理

QAM 信号调制时，首先将两路独立的二进制基带信号  $x'(t)$  和  $y'(t)$ ，转换为 2 通道的  $M$  进制基带信号  $x(t)$  和  $y(t)$ ，然后用同频正交载波分别调制，两路合成输出即可。图 5-9 给出了产生多进制 QAM 信号的数学模型。

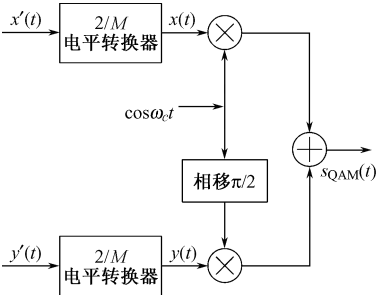


图 5-9 QAM 信号产生的数学模型

QAM 信号解调一般采取正交相干解调的方法，其数学模型如图 5-10 所示。

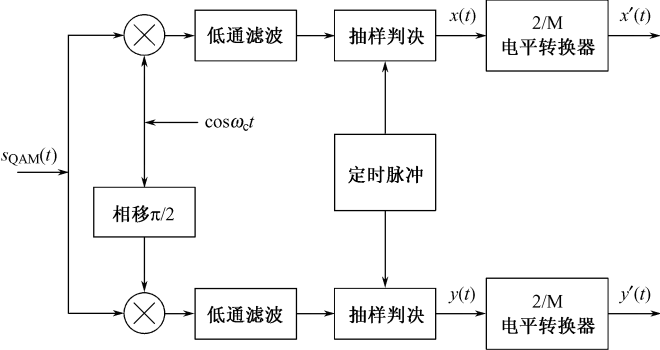


图 5-10 QAM 信号正交相干解调



解调器首先对收到的 QAM 信号进行正交相干解调。低通滤波器 LPF 滤除乘法器产生的高频分量，经抽样判决可恢复出  $M$  进制电平信号  $x(t)$  和  $y(t)$ 。

因为  $x_k$  和  $y_k$  取值一般为  $\pm 1, \pm 3, \cdots, \pm(M-1)$ ，所以判决电平应设在信号电平间隔的中点，即  $U_d=0, \pm 2, \pm 4, \cdots, \pm(M-2)$ 。

根据多进制码元与二进制码元之间的关系，经  $M/2$  转换，可将  $M$  电平信号转换为二进制基带信号  $x'(t)$  和  $y'(t)$ 。

### 3. QAM 与 PSK 性能比较

对于相同状态数的多进值数字调制,QAM抗噪性能优于PSK,下面就以16QAM和16PSK性能进行比较。图 5-11 分别给出了这两种信号的星座图。

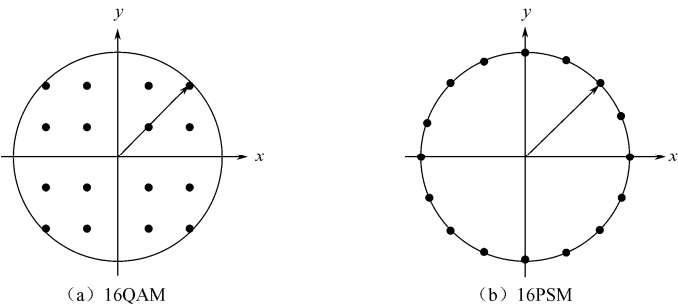


图 5-11 16QAM 和 16PSK 信号星座图

设 16QAM 和 16PSK 信号最大振幅为  $A$ ，则相邻矢量端点的距离分别为

$$\left. \begin{aligned} d_{16\text{PSK}} &= 2A \cdot \sin \pi/16 \approx 0.39A \\ d_{16\text{QAM}} &= \sqrt{2} A/2 \approx 0.47A \end{aligned} \right\} \tag{5-29}$$

相邻矢量端点的距离越大，其抗干扰能力越强。从式 (5-29) 可以看出， $d_{16\text{PSK}} < d_{16\text{QAM}}$ ，因此，在最大功率（振幅）相等的条件下，16QAM 抗噪性能优于 16PSK。同样还可以证明，在平均功率相等的条件下，16QAM 抗噪性能仍然优于 16PSK。

### 5.3.2 可变速率正交振幅调制（VR-QAM）

可变速率正交振幅调制（Variable Rate QAM, VR-QAM）是一种在正交振幅调制基础上，保持数据传输比特率不变条件下，根据衰落信道特性，自动降低码元速率（增加调制电平数），或增加码元速率（或减少调制电平数），即改变信号星座图，实现调制的一种方式，由于其码元传输速率是变化的，所以称这种方式为可变速率正交振幅调制。

#### 1. VR-QAM 的星形星座图表示

VR-QAM 采用星形星座图，如图 5-12 所示，比方形星座图有较少的振幅和相位。例如，16QAM 星形星座图 5-12 (d) 与 16QAM 方形星座图 5-11 (a) 对应，其振幅圆环由 3 个减少为 2 个，相位由 12 个减少为 8 个，这有利于接收端的自动增益控制和载波相位跟踪，因此更适合于衰落信道。

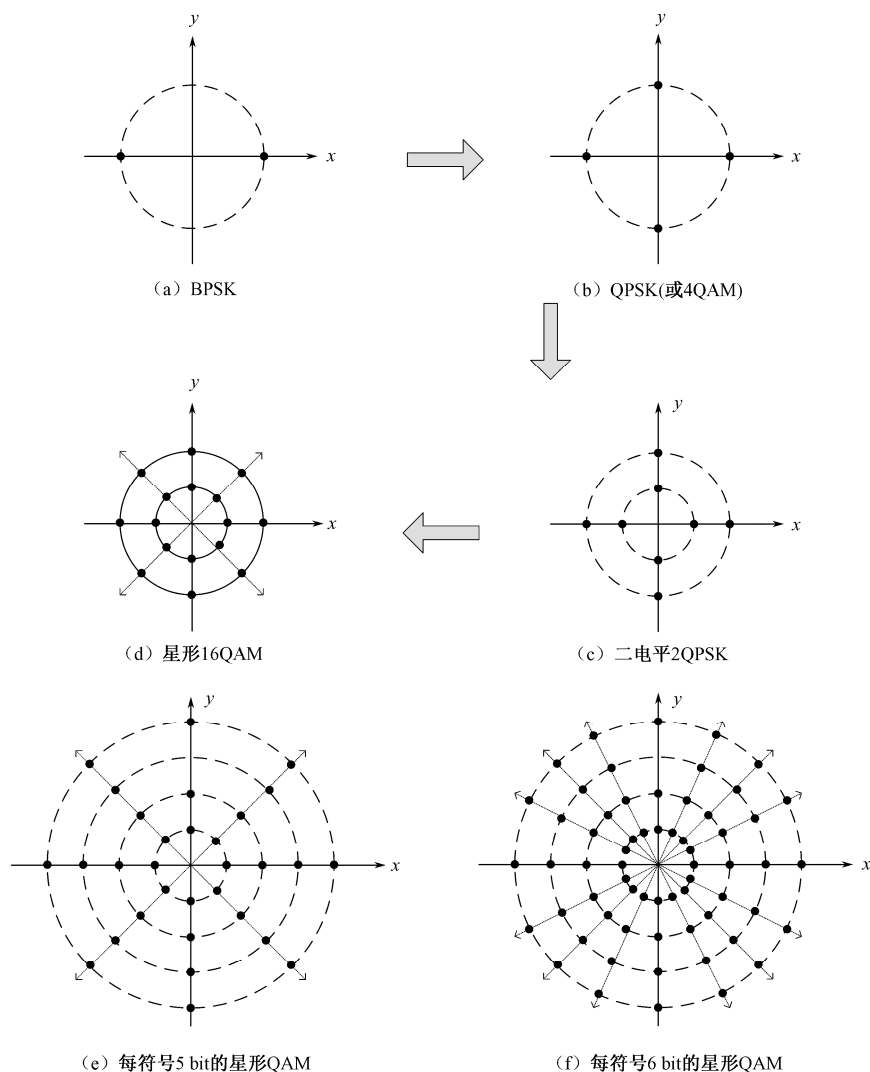


图 5-12 VR-QAM 的星形星座图

在调制过程中，调制电平数的增加或减少，即每码元包含比特数的增加或减少，是采用加倍星座图幅度圆环数目和加倍相位点数目交替进行的。从每符号 1 bit 的 BPSK 信号开始，对于每符号 2 bit，加倍相位点数目得到 QPSK，对于每符号 3 bit，加倍幅度圆环数目，得到二电平 QPSK，对于每符号 4 bit，加倍相位点数目得到 16 值星形 QAM，对于每符号 5 bit，加倍幅度圆环数目得到 32 值星形 QAM，对于每符号 6 bit，加倍相位点数目得到 64 值星形 QAM，共有 4 个幅度环，每个幅度环上有 16 个相位点。

VR-QAM 信号的调制星座图，由于不是恒定的，所以必须在发送信号中包含这方面信息，以便接收机能够选择解调方案，正确地解调信号。

## 2. VR-QAM 的信道信息获得及数据帧结构

为了获得信道条件好坏的信息，VR-QAM 应该工作在双工传输系统中。适合于可变速率

调制方式的双工方式为时分双工（TDD）。时分双工的收、发用同一个频率，但占用不同的时间。如果收、发分别占用 TDD 帧的一半时间，且选择合适的帧长（为使上行和下行信道衰落状况接近），就可根据收到的信号进行信道评估，从而确定即将发出数据应采用的 QAM 电平数。图 5-13 是移动通信基地台（BS）和移动台（MS）的 VR-QAM 的数据帧结构。

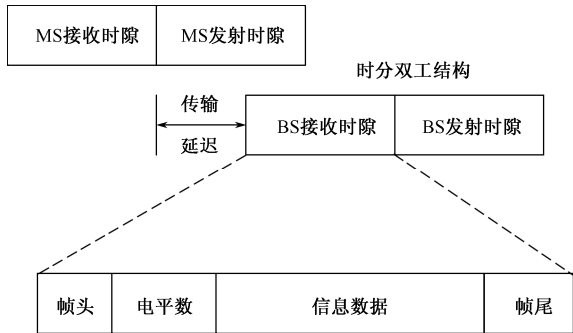


图 5-13 VR-QAM 的数据帧结构

MS 和 BS 都是先接收，再发射。一个时隙包括帧头、电平数、信息数据和帧尾，各段作用为：

- 帧头：包含同步序列和训练序列，用于提取同步信息；
- 电平数：包含当前数据块中 QAM 调制电平数信息，接收端利用该信息选择合适的解调方式，此段数据采用 QPSK 调制，信息占两个码元，重复 3 次。
- 信息数据：包含采用 VR-QAM 要传输的数据；
- 帧尾：是时分方式的保护时间。

### 3. VR-QAM 的性能

移动通信中 VR-QAM 的性能与移动台的速度、数据块长、同道干扰的强弱有关。研究表明有如下结论：

- （1）若 VR-QAM 的平均速率与固定 QAM 速率相同，则 VR-QAM 的性能比固定 QAM 的性能有明显改善，并且在移动台低速移动的情况下，有更加明显的改善。
- （2）块（帧）的长度选择在 VR-QAM 中很重要，块长过长，块内采用电平数不一定适合信道变化，块长过短，则需传送更多的电平变化信息，从而降低总的信息通过率。一般来说，当移动台移动速度较慢时，块长较长，移动台移动速度较快时，长的块长会使性能下降。
- （3）对移动通信中的同道干扰，VR-QAM 比固定 QAM 性能好。

## 5.4 正交频分复用（OFDM）

### 5.4.1 OFDM 基本原理

正交频分复用（Orthogonal Frequency Division Multiplexing，OFDM）是多载波通信的一种，与一般的频分复用（FDM）技术不同，在 OFDM 系统中各子信道在时间上互相正交，在频率上互相重叠。

传统的多载波传输系统，是将高速率的信息数据流经串/并变换，分割为若干路低速率并行数据流，每路低速率数据采用一个独立载波调制，采用频分复用调制方式叠加在一起构成发送信号，其子载波的频率设置如图 5-14 所示。接收端，首先要采用滤波器组，进行各子带的分离，然后解调各路低速率数据，再经并/串变换，恢复出高速率的信息数据流。这种方案的优点是实现简单、直接；缺点是频率利用率低，各子载波之间要保留有保护带，而且多个滤波器的实现也有不少困难。

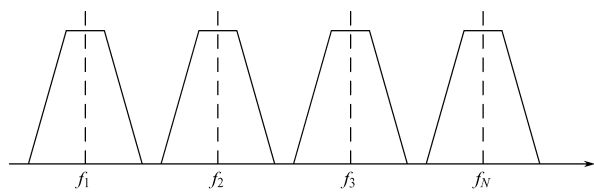


图 5-14 传统的子载波的频率设置

OFDM 是一种特殊的多载波传送方案，各子载波频谱有 1/2 重叠，但保持相互正交，接收端通过相关解调技术分离出各子载波，同时消除码间干扰的影响。这种方案的优点是可以避免使用滤波器组，同时提高频率利用率近 1 倍。图 5-15 为 OFDM 调制原理框图。

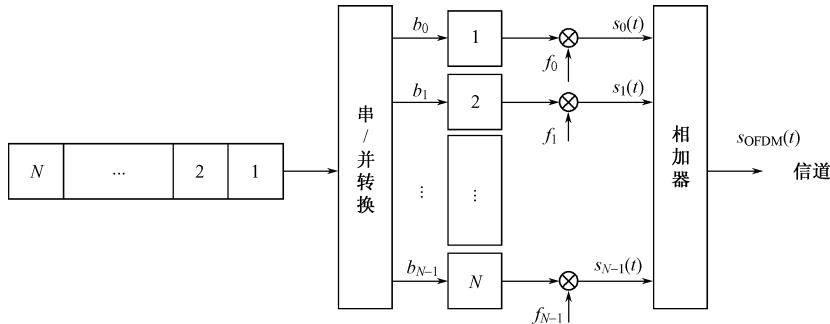


图 5-15 OFDM 调制原理框图

$N$  个待发送的串行数据经串/并变换后，得到周期为  $T_b$  的  $N$  路并行码，码型选用双极性 NRZ 矩形脉冲，经  $N$  个子载波分别对  $N$  路并行码进行调制，相加后得到波形

$$s_{\text{OFDM}}(t) = \sum_{k=0}^{N-1} b_k \cdot \cos \omega_k t \tag{5-30}$$

式中， $b_k$  为第  $k$  路并行码； $\omega_k$  为第  $k$  路子载波角频率。

为了保证  $N$  个子载波相互正交，也就是在信道传输符号的持续时间  $T_b$  内，它们的乘积的积分值为零。由三角函数的正交性

$$\int_0^{T_b} \cos\left(2\pi \frac{m \cdot t}{T_b}\right) \cdot \cos\left(2\pi \frac{n \cdot t}{T_b}\right) \cdot dt = \begin{cases} 0 & m \neq n \\ \pi & m = n \end{cases} \quad (m, n = 1, 2, \dots) \tag{5-31}$$

可知，子载波频率间隔应为

$$\left. \begin{aligned} \Delta f &= f_k - f_{k-1} = 1/T_b \\ f_k &= f_0 + k/T_b \\ k &= 1, 2, \dots, N-1 \end{aligned} \right\} \tag{5-32}$$

式中,  $f_0$  为最低子载波角频率。

由式 (5-32) 可知, 当 OFDM 信号各子路的码型选用 NRZ 矩形脉冲时, 每路信号频谱形式为  $\text{Sa}(\omega T_b/2)$  函数, 则相邻信号频谱之间有  $1/2$  重叠, 图 5-16 为 OFDM 频谱结构。

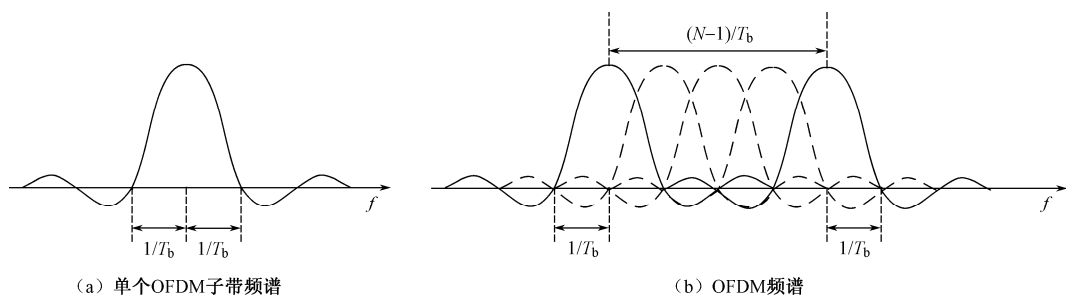


图 5-16 OFDM 频谱结构

忽略旁瓣的功率, OFDM 信号的频谱宽度为

$$B_{\text{OFDM}} = (N-1) \cdot \frac{1}{T_b} + \frac{2}{T_b} = \frac{N+1}{T_b} \quad (5-33)$$

因信道在  $T_b$  时间内, 能传输  $N$  个并行码元, 则码元速率为  $R_B = N/T_b$ , 对应频带利用率为

$$\eta_{\text{OFDM}} = \frac{R_B}{B_{\text{OFDM}}} = \frac{N}{N+1} \approx 1 \text{ bit} \cdot \text{s}^{-1}/\text{Hz} \quad (5-34)$$

在接收端, 对  $s_k(t)$  用频率  $f_k(k=0, 1, \dots, N-1)$  的正弦载波, 在  $[0, T_b]$  内进行相关运算, 就可得到各子载波携带的信息  $b_k$ , 然后通过并/串变换, 恢复出发送的二进制数据序列。图 5-17 为 OFDM 信号的解调原理框图。

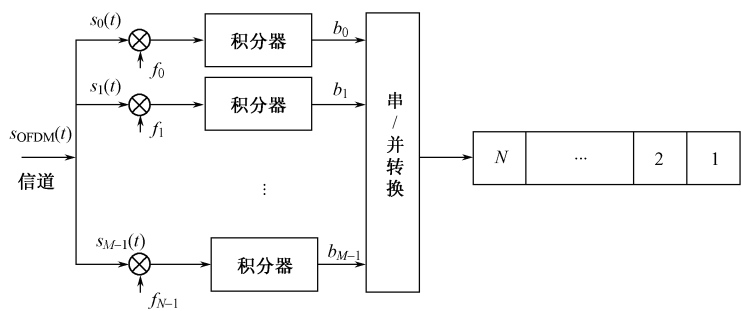


图 5-17 OFDM 信号的解调原理框图

### 5.4.2 用离散傅里叶变换实现 OFDM

OFDM 信号的调制与解调, 按照图 5-15 和图 5-18 给出的实现的方法, 所需要的设备非常复杂, 特别是当  $N$  很大时, 需要大量的正弦波发生器、调制器和相关解调器等设备, 费用非常昂贵。但是随着信号处理理论和技术的发展, 到 20 世纪 80 年代, 快速傅里叶变换 (Fast Fourier Transform, FFT) 算法和器件日趋成熟, 人们提出采用离散傅里叶反变换 (IDFT) 来实现多个载波的调制, 利用离散傅里叶变换 (DFT) 来实现解调, 可以极大地降低 OFDM 系统的复杂度和成本, 从而使得 OFDM 技术更趋于实用化。其具体原理如下:

式 (5-30) 可以改写为复数形式

$$s_{\text{OFDM}}(t) = \text{Re} \left[ \sum_{k=0}^{N-1} b_k \cdot e^{j\omega_k t} \right] = \text{Re} \left[ \sum_{k=0}^{N-1} b_k \cdot \exp(j\omega_k t) \right] \quad (5-35)$$

如果对 OFDM 信号在  $[0, T_b]$  内进行  $N$  点离散化, 其抽样间隔  $T = T_b / N$ , 则抽样时刻  $t = nT$  的 OFDM 信号为

$$\begin{aligned} s_{\text{OFDM}}(nT) &= \text{Re} \left[ \sum_{k=0}^{N-1} d(k) \cdot \exp(j\omega_k nT) \right] \\ &= \text{Re} \left[ \sum_{k=0}^{N-1} d(k) \cdot \exp(j\omega_k nT_b / N) \right] \end{aligned} \quad (5-36)$$

式中, 离散序列  $d(k) = b_k$ ,  $k = 0, 1, 2, \dots, N-1$ 。

由于 OFDM 信号的产生首先是在基带实现, 然后通过上变频产生输出信号。因此, 处理时为了方便, 令最低子载波角频率  $\omega_0$  为零, 根据式 (5-32) 则

$$\omega_k = 2k\pi / T_b \quad (5-37)$$

将式 (5-37) 代入式 (5-36) 则

$$s_{\text{OFDM}}(nT) = \text{Re} \left[ \sum_{k=0}^{N-1} d(k) \cdot \exp(j2\pi \cdot k \cdot n / N) \right] \quad (5-38)$$

由于等号右边与  $T$  无关, 则可以写为

$$s_{\text{OFDM}}(n) = \text{Re} \left[ \sum_{k=0}^{N-1} d(k) \cdot \exp(j2\pi \cdot k \cdot n / N) \right] \quad (5-39)$$

考虑到长度为  $N$  的序列  $x(n)$ , 其  $N$  点离散傅里叶变换 (DFT) 可以写为

$$X(k) = \sum_{n=0}^{N-1} x(n) \cdot \exp(-j2\pi \cdot k \cdot n / N) \quad k = 0, 1, \dots, N-1 \quad (5-40)$$

相应  $N$  点离散傅里叶反变换 (IDFT) 可以写为

$$x(n) = \frac{1}{N} \sum_{k=0}^{N-1} X(k) \cdot \exp(j2\pi \cdot k \cdot n / N) \quad n = 0, 1, \dots, N-1 \quad (5-41)$$

比较式 (5-39) 和式 (5-41) 可以看出, OFDM 信号的产生可以用离散傅里叶反变换实现, OFDM 信号的解调可以用离散傅里叶正变换实现, 工程上可以采用 FFT 技术, 图 5-18 给出了用 DFT 实现 OFDM 的原理。

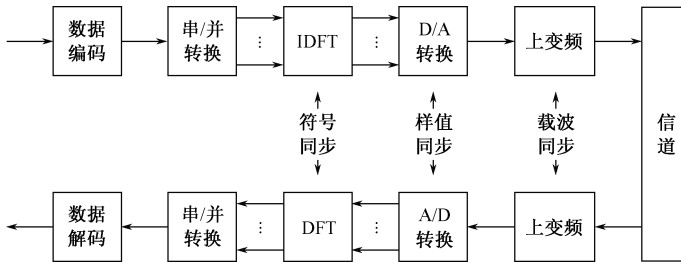


图 5-18 用 DFT 实现 OFDM 的原理框图

在发送端, 输入的二进制数据序列先进行串/并变换, 得到  $N$  路并行码, 经 IDFT 变换得 OFDM 信号数据流各离散分量, 送入 D/A 变换形成双极性多电平方波, 再经上变频调制最后形成 OFDM

信号发送出去。在接收端，OFDM 信号的解调过程是其调制的逆过程，这里不再赘述。

### 5.4.3 OFDM 的优缺点

OFDM 的主要优点：

(1) 可以有效地解决多径问题；对一定的延迟扩展，实现难度远远低于使用均衡器的单载波系统。

(2) 在相对低时变信号中，根据特定子载波的信噪比来调整每个子载波的数据传输速率，可以有效地增加容量。

(3) 可以很好地抵抗窄带干扰，这是因为 OFDM 一是子带很窄，远小于信道相干带宽；二是在传输中，只有少量子载波受伤害，易于采用纠错加以校正，所以可以提高抗频率选择性衰落和窄带干扰的能力。这些干扰只影响到一部分子载波。

OFDM 的缺点：

(1) OFDM 对频率偏移和相位噪声更加敏感。

(2) OFDM 需要相对更高的峰值平均功率比，这会降低射频放大器的功效。

## 思考与练习

5-1 何谓调制？何谓解调？何谓无线调制与解调？

5-2 何谓线性调制？什么是恒包络调制？

5-3 为什么要改进 4PSK 调制方案，OQPSK 和  $\pi/4$ -QPSK 是怎样改进的，改进的效果如何？

5-4 MSK 是一种怎样的调制方案？它的优点有哪些？适合于哪种通信系统？

5-5 画图说明 MSK 调制器与相干解调原理框图。

5-6 GMSK 与 MSK 比较，改进点在哪？调制系统方框与 MSK 有哪些不同？

5-7 画图说明 QAM 调制器与相干解调原理框图，并与 PSK 性能比较。

5-8 什么是可变速率正交振幅调制？

5-9 画图 OFDM 调制器与相干解调原理框图，并进行说明。

5-10 画出用 DFT 实现 OFDM 的原理框图，说明 OFDM 的主要优缺点。

# 第 6 章 抗干扰通信与通信干扰技术

现代战争中信息技术正起着越来越大的作用，制信息权和制空权、制海权一样，是战争胜负的关键，在夺取制信息权的信息战中，军事通信技术的发展始终走在民用技术的前面，并“以通信为中心”向“以通信对抗为中心”的转变，就逐渐成为了一种特殊的战斗形态，这就是通信电子战，也被称为通信对抗。

根据抗干扰通信与通信干扰的技术特点，从频域角度，有跳频、扩频技术等；从空域角度，有智能天线技术等；从时域角度，有信道编码技术等，本章主要从频域角度研究、探讨。

## 6.1 引言

### 6.1.1 通信对抗基本概念

#### 1. 通信对抗的含义

在无线电通信过程中，通信发射机向空间辐射载有信息的无线电信号，而通信接收机则在复杂的电磁环境中检测出有用的信息，这种开放式发射和接收通信信号的特点是实施无线电通信对抗的物理基础。

图 6-1 给出了一个简单的通信对抗基本原理示意图，其中红方是执行通信联络任务的一方，蓝方是对红方进行通信干扰的一方。在这个对抗中，蓝方的目的是根据红方通信样式，释放合适的干扰信号到达红方通信接收机，使之无法实现正常接收，从而破坏红方的通信联络；而红方的目的是不管蓝方如何干扰，都要保证待传信息能够在接收机中被正确接收和解调。由此可见，通信对抗双方就是一对“矛”和“盾”的关系，这个“矛”就是在通信侦察等措施支援下的通信干扰，这个“盾”就是抗干扰通信。本章的内容就是在阐述通信对抗基本概念的基础上，重点围绕通信领域的这对“矛”和“盾”来展开讨论。

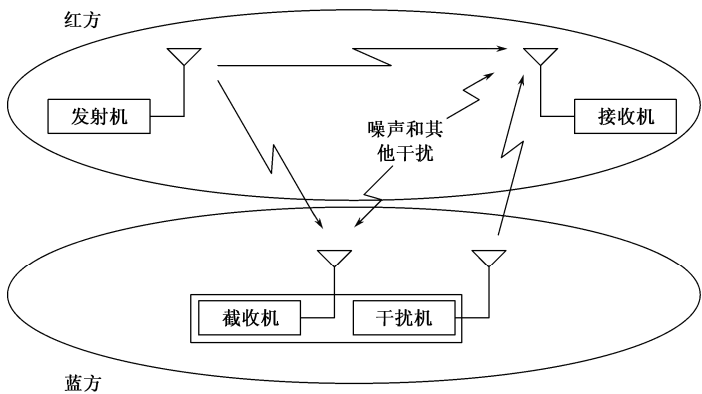


图 6-1 通信对抗原理示意图



2. 通信对抗的基本内容

通信对抗的基本内容包括通信电子侦察、通信电子进攻和通信电子防御三部分。

1) 通信电子侦察

通信电子侦察就是通过使用通信电子侦察接收设备和无线电测向设备，对敌无线电通信设备所发射的通信信号进行搜索、截获、测量分析和对敌方通信信号进行测向定位，以获取信号频率、电平、通信方式、调制样式和电台位置信息，对其侦听判别，以确定信号的属性。

通信电子侦察的最大特征就是无源工作，即通信侦察本身不向外发射信号，它的任务就是接收来自空间的各种电磁信号，并从中提取出有用的信号并进行参数估计。

2) 通信电子进攻

通信电子进攻是在通信领域实施的电子攻击，是破坏、影响、削弱敌方通信系统作战使用效能的主要手段。按照进攻手段的不同，通信电子进攻可以进一步分为通信干扰和通信摧毁两项主要内容。

通信干扰是针对无线电通信系统的接收机所实施的人为电磁频谱干扰。

通信摧毁是采用化学能、定向能以及激光等武器对敌方通信设备和系统实施摧毁，是现代通信电子进攻的重要手段。

3) 通信电子防御

通信电子防御，就是在通信对抗斗争中，为保证己方通信设备及其系统正常发挥效能而采取的措施与行动的统称。这就要求对抗双方使用各种手段和方式，以使己方在对抗中处于有利地位。通信电子防御包括通信反侦察、通信反干扰及通信反摧毁等内容。

为简单起见，图 6-2 给出了通信对抗的基本内容，其中阴影部分是本章重点讨论内容，其余内容可参看相关文献，本章不再赘述。

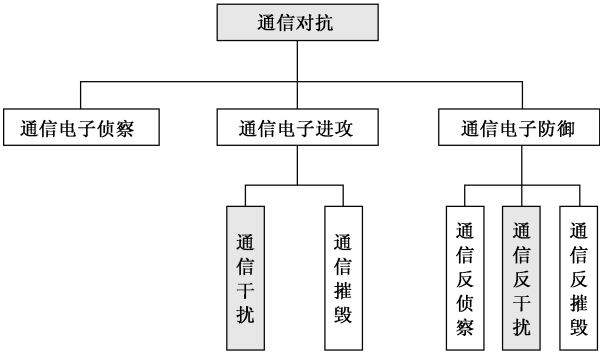


图 6-2 通信对抗的基本内容

6.1.2 抗干扰通信的基本概念

所谓抗干扰通信，就是在各种干扰条件或复杂电磁环境中保证通信正常进行的各种技术和战术措施的总称。

常用的抗干扰通信技术有两大类，一类是基于扩展频谱的抗干扰通信技术，另一类是基于非扩展频谱的抗干扰通信技术。

所谓扩展频谱（Spread Spectrum, SS），就是将信息带宽进行扩展传输的一种抗干扰通信手段。根据频谱扩展的方式不同，扩谱又可以分为直接序列扩谱（Direct Sequence Spread Spectrum, DSSS, 简称直扩或 DS），跳频扩谱（Frequency Hopping Spread Spectrum, FHSS, 简称跳频或 FH），跳时扩谱（Time Hopping Spread Spectrum, THSS, 简称跳时或 TH），调频扩谱（Chirp）和混合扩谱（Hybrid Spread Spectrum）等。

基于非扩展频谱的抗干扰通信体制主要是指不通过对信号进行频谱扩展而实现抗干扰的技术方法总称。目前常用的方法主要有自适应滤波、干扰抵消、自适应频率选择、捷变频、功率自动调整、自适应天线调零、智能天线、信号冗余、分集接收、信号交织和信号猝发等，同样属于抗干扰通信的研究范畴，且近年来该领域的研究逐渐升温，成为抗干扰通信的研究热点。

和基于扩展频谱的抗干扰通信体制相比，基于非扩展频谱的抗干扰方法所涵盖的范围更广，所涉及的知识也更多。通过二者比较不难发现，前者主要是在频率域、时间域以及速度域上来考虑信号的抗干扰问题，而后者除了涉及上述三个领域外，还将在功率域、空间域、变换域以及网络域等方面下功夫。

虽然抗干扰通信的方法很多，但从本质上来讲，所有技术方法的最终目的只有一个，就是提高通信系统接收端的有效信噪比（Signal to Noise and Interference Ratio, SNIR），从而保证接收机能够正常地实现对有用信号的正确接收。

### 6.1.3 通信干扰基本概念

所谓通信干扰，就是采用一切手段来阻止敌方对电子通信、降低或破坏敌方通信电子设备的作战使用效能。

对于无线电通信中存在的干扰，人们在日常生活中并不陌生。冲击钻旋转时会引起电视接收机出现雪花，放在计算机旁的手机会时不时地导致显示器闪动，这些就是无线电干扰的现象，只不过这些干扰是无目的的且时有时无的。和自然界中的上述干扰有本质区别的是，电子战中的无线电通信干扰特指一种有目的的人为破坏。这种人为破坏之所以能够实现，主要基于两大物理基础，分别是无线电信道的开放性和无线电波传输的透明性。

通信干扰的基本方法是将干扰信号随同敌方所期望的通信信号一起送到敌方接收机中。当敌方接收机中的干扰信号强度达到足以使敌方无法从接收到的信号中提取有用信息时，干扰就是有效的。因此，通信干扰得以应用的最基本对象就是干扰对方的接收机，而不是发射机。要达到干扰有效的目的，干扰机必须在天线发射方向、发送信号强度、干扰距离以及传播条件等方面进行精心考虑。

## 6.2 抗干扰通信技术

### 6.2.1 抗干扰通信的理论基础

前面已经谈到，扩展频谱（简称扩谱）是目前军事通信中广泛采用的基本抗干扰技术体制，很多经典著作对其基本理论进行了详细的描述，本节也以扩谱通信体制为主，来论述抗

干扰通信技术所涉及的理论基础，为后续讨论奠定理论基础。

### 1. 香农公式及其工程意义

香农公式是一个被广泛公认的通信理论基础和研究依据，也是近代信息论的基础。扩谱通信的理论基础是著名的香农（Shannon）公式，即对于高斯白噪声信道有

$$C = B \log_2(1 + S/N) \tag{6-1}$$

式（6-1）中， $C$  为信道容量，单位为 bit/s； $B$  为传输信息所用的带宽，单位为 Hz； $N$  为噪声平均功率， $S$  为信号平均功率， $S/N$  为信号与噪声的功率之比。

由于噪声平均功率  $N$  与系统带宽  $B$  有关，假设单边噪声功率谱密度为  $n_0$ ，则噪声平均功率  $N = n_0 B$ 。因此，香农公式的另一种表达形式为

$$C = B \log_2 \left( 1 + \frac{S}{n_0 B} \right) \tag{6-2}$$

由式（6-2）可见，信道容量  $C$  与“三要素” $B$ 、 $S$ 、 $n_0$  有关，只要这三个要素确定，信道容量  $C$  也就随之确定。

从香农公式可以看出，在单边噪声功率谱密度  $n_0$  一定的条件下，一个给定的信道容量可以通过增加带宽  $B$  而减小信号功率  $S$  的办法实现，也可以通过增加信号功率而减小带宽的办法实现。这就是说，信道容量可以通过带宽与信号功率或信噪比的互换而保持不变。也可以说，分别通过增加信号功率  $S$  和带宽  $B$  都可以提高信道容量  $C$ 。但是，哪种方式的效果更好呢？由式（6-1）并参照对数函数关系，在大信噪比条件下（即  $S/N \gg 1$ ），式（6-2）可近似写成

$$C = B \log_2(S/N) \tag{6-3}$$

此时若信号功率  $S$  不变，信道容量  $C$  与带宽  $B$  近似成线性关系，上升速度较快；若带宽  $B$  不变，信道容量  $C$  与信号功率  $S$  近似成对数关系，上升速度较缓慢。

同样，由式（6-3）及对数特性，在小信噪比条件下（即  $S/N \ll 1$ ），信道容量  $C$  与带宽  $B$  仍近似成线性关系，而信道容量  $C$  与信号功率  $S$  虽然仍成对数关系，但此时变化的斜率较大，也近似于直线的变化斜率。由此可见，在大信噪比条件下，若采用增加带宽，去换取功率的减小，只要增加较小的带宽就可以节省较大的功率，或者说以带宽换功率的效果更好；而在小信噪比条件下，两种方式的效果相当，理论分析和工程实践可以证明这一点。

理论上还可以证明，在具有极限信息传输速率的理想系统中，输出信噪比随着带宽的增加而按指数规律增加。也就是说，增加带宽可以明显地改善输出信噪比。

根据带宽与功率互换的这一原理，应该尽可能扩展信号的传输带宽，以提高系统的输出信噪比，这就是扩展频谱通信。比如，跳频通信射频覆盖的带宽比信号的原始带宽大得多，直扩后的信号带宽比直扩前的信号带宽大得多。

### 2. 处理增益

任何一个（扩谱）处理单元或系统的输出与输入信噪比之间的比值称为其（扩谱）处理增益  $G_p$ ，定义为

$$G_p = \frac{(S/N)_0}{(S/N)_i} = \frac{S_o/n_0 B_o}{S_i/n_0 B_i} = B_i/B_o \tag{6-4}$$

通常以分贝形式表示为

$$G_p(\text{dB}) = 10 \lg(B_i/B_o) \tag{6-5}$$

式(6-4)中,  $(S/N)_o$  为系统输出信噪比,  $(S/N)_i$  为系统输入信噪比,  $n_0$  为高斯噪声功率谱密度,  $B_i$  为系统的输入信号带宽,  $B_o$  为系统的输出信号带宽。一般来说, 输入信号功率  $S_i$  若在处理过程中没有损失, 它应该和输出信号功率  $S_o$  相等。

$G_p$  表明(扩谱)系统前后信噪比改善程度, 体现(扩谱)系统有用信号增强, 干扰受到抑制的能力。 $G_p$  值越大, (扩谱)系统的抗干扰能力越强。

对于跳频体制,  $B_i$  即为射频跳变时所能覆盖的全部射频带宽, 也就是从最低频率到最高频率所覆盖的全部射频频率范围(通常称之为跳频带宽);  $B_o$  即为跳频的瞬时带宽, 也就是经跳频压缩和中频调制后的带宽或经单点射频搬移后的射频瞬时带宽, 该瞬时带宽是窄带的, 略宽于相同数据速率信号的定频通信中频调制后的带宽。可见, 跳频体制的射频瞬时带宽与其跳变覆盖的全部射频带宽是不一样的。

对于直扩体制,  $B_i$  即为直扩后的带宽, 也就是信码经伪码序列调制后的中频带宽或经射频搬移后的信号带宽;  $B_o$  即为直扩前的原信号带宽。可见, 直扩体制的射频瞬时带宽与直扩中频调制带宽是一样的, 但与原信息带宽是不一样的。

对于直扩/跳频体制,  $B_i$  即为射频跳变时所能覆盖的全部射频带宽,  $B_o$  即为原信息的信号带宽。在相同信息速率和相同直扩方式条件下, 与无直扩的纯跳频体制相比, 直扩/跳频体制的瞬时带宽即为直扩带宽, 对频率资源的需求也增加了; 与单频点的纯直扩体制相比, 直扩/跳频体制的载频数量和射频覆盖范围增加了, 但对直扩瞬时带宽提出了限制。由于直扩/跳频体制也是扩谱体制, 其处理增益  $G_{DH}$  仍可按式(6-4)和式(6-5)计算, 很容易得出直扩/跳频处理增益  $G_{DH}$  与直扩处理增益  $G_{DS}$  和跳频处理增益  $G_{FH}$  的关系为

$$G_{DH} = G_{DS} \cdot G_{FH} \quad (6-6)$$

以分贝表示为

$$G_{DH}(\text{dB}) = 10\lg G_{DS} + 10\lg G_{FH} \quad (6-7)$$

虽然直扩/跳频体制的处理增益等于直扩处理增益与跳频处理增益之和, 但并不是说直扩/跳频处理增益就大大增加了。实际上, 只要射频工作带宽没有增加, 直扩/跳频处理增益也不会增加, 因为此时瞬时带宽增加了, 可用频点减少了, 并且此时直扩带宽也不能像纯直扩那样宽。在射频带宽资源相等的情况下, 实际的直扩/跳频体制处理增益与纯直扩或纯跳频处理增益相当。尽管如此, 直扩/跳频体制仍具有较多的优势, 将在有关参考书中介绍。

扩谱处理增益表示了系统解扩前后信噪比改善的程度和敌方干扰扩谱系统所要付出的理论上的代价, 是系统抗干扰能力的重要指标, 但仅是理论上的抗干扰能力。

### 3. 干扰容限

本节在讨论经典干扰容限概念及其工程意义的基础上, 重点讨论直扩和跳频的干扰容限及其区别。

现在的问题是, 扩谱系统到底能容忍多大的干扰还能正常工作。由于处理增益仅是一种理论上的概念或理论值, 用处理增益还不能完全解释这个实际的问题。根据前人的研究成果, 将扩谱通信系统能维持点对点正常工作(满足正常解调要求的最小输出信噪比)的实际抗干扰能力定义为干扰容限  $M_j$ , 其表达式为

$$M_j = G_p - [L_s + (S/N)_{out}] \quad (6-8)$$

与式(6-1)类似, 式(6-8)的推导过程与调制类型和其他信道参数无关。但对于实际系统, 式(6-8)中各参量值与系统的调制类型和相应的信道参数有关。在式(6-8)中,  $(S/N)_{out}$  为接收机解调输出端所需的最小信噪比;  $L_s$  为扩谱系统解扩/解调的固有处理损耗, 它是由扩

谱信号处理以及工程实现中的误差对信号造成的损伤而引起的。可见，实际中希望  $L_S$  和  $(S/N)_{out}$  越小越好，干扰容限一般小于处理增益。根据工程经验， $L_S$  一般为 1~2.5 dB，最大不超过 3 dB， $[L_S + (S/N)_{out}]$ （即干扰容限与处理增益的差值）一般为 5 dB 数量级，对于不同的技术方案和工程实现水平， $L_S$  和  $(S/N)_{out}$  的具体值有所不同，在设备研制和验收中应对此进行界定和考核。对于一个实际的扩谱系统，上式右边的各参数都是确定的值，即干扰容限也是一个确定的值，无论是敌方干扰还是非敌方干扰都将消耗干扰容限，所以在实际使用中应尽可能减少或避免非敌方干扰，以发挥扩谱系统抗敌方干扰的潜力。

顺便指出的一点是，这里提到的术语是干扰容限，不能将其称之为“抗干扰容限”，在指标界定中经常出现误用，应引起注意。

基于干扰容限的物理意义，在技术方案制定和信道机设计中，要着力提高扩谱系统的干扰容限。

上式表明，干扰容限与扩谱处理增益、系统的固有处理损耗和输出端所需的最小信噪比三个因素有关，扩谱处理增益越大，系统的固有处理损耗和解调所需的最小信噪比越小，干扰容限就越大。所以，应尽量提高处理增益，降低系统的固有处理损耗和解调所需的最小信噪比。系统的处理增益主要与信息速率、频率资源、扩谱解扩方式等因素有关，系统的固有处理损耗和解调所需最小信噪比主要与扩谱解扩方式、交织与纠错方式、调制解调性能、自适应处理、信号损伤、同步性能、时钟精度、器件稳定性、弱信号检测能力、接收机灵敏度等指标有关。这些都是提高干扰容限和系统基本性能的切入点。同时上式也说明，尽管处理增益与系统抗干扰能力有直接的关系，但不能完全表明系统的抗干扰能力，还与其他因素有关。

以上讨论的干扰容限的基本概念（即干扰容限表明了系统维持点对点正常通信的实际抗干扰能力）对于几种扩谱体制都是适用的，但对于不同的扩谱体制，干扰容限的表现形式则不尽相同。

## 6.2.2 扩谱抗干扰通信技术

### 1. 直接序列扩谱（DSSS）

直接序列扩展频谱系统，又称“平均”系统或伪噪声系统，它是目前应用较广泛的一种扩展频谱系统。为大家所熟知的 GPS（Global Position System）就是一种典型的直扩系统，在国外已获得成功的空间探测器“喷气推进实验室”（JPL）测距技术门也采用了直接序列扩谱调制，目前 DS/SS 还用于各种测试系统和实验测试设备，预计它将会被其他领域所采用。

图 6-3 给出了一种典型的 BPSK/DSSS 系统原理方框图。图中发送的数据经过编码器后，首先进行 BPSK 调制，然后用产生的伪随机序列对 BPSK 信号进行直扩调制，扩谱后的宽带信号经功放后由天线发射出去。接收端接收到的信号经过前端射频放大后，用本地伪随机序列对直扩信号完成“逆扩谱调制”，然后信号通过窄带带通滤波器，之后与本地载波相乘去载波，再经过低通滤波、积分抽样后，送至数据判决器，恢复出数据。

图 6-3 中虚线部分分别完成扩谱调制与解扩的作用。在该模型中，扩谱伪随机序列和信息同样采用了 BPSK 调制方式，扩谱的调制是通过直接对载波的调制来实现的。直扩信号可以表示为：

$$s(t) = \sqrt{2P} d(t)c(t)\cos(\omega_0 t + \varphi_0) \tag{6-9}$$

其中， $P$  为直扩信号的平均功率； $d(t)$  是双极性单位功率的基带数据信号，取值为  $\pm 1$ ； $c(t)$  是双极性单位功率的伪随机序列信号； $\omega_0$  是载波角频率； $\varphi_0$  是载频的初相。

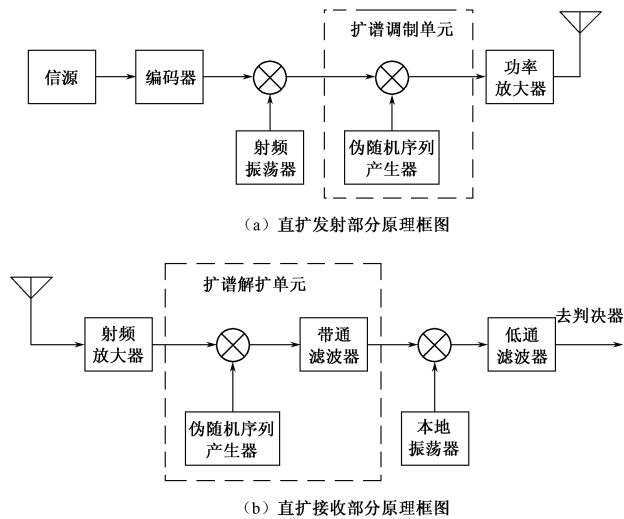


图 6-3 直扩系统原理方框图

由式 (6-9) 可见， $d(t)$  完成的是信息的调制，而  $c(t)$  完成了直扩调制。在信号格式不变的情况下，显然扩谱调制和信息调制的顺序是可以改变的。并且，双极性序列的相乘对应二元序列的异或运算。因此，直扩调制也可以用数字电路完成。当扩谱伪随机序列与信息序列的时钟是同步的时，经过数字处理后输出的高速基带序列的速率与伪随机序列的速率相同，即调制后带宽取决于伪随机序列的速率。

图 6-4 给出了系统工作的各部分波形示意图，其中 (a) 是发射机的工作波形，(b) 是接收机的工作波形，图 6-4 中的基带数据和伪随机序列信号都是单位功率的双极性信号。由图可见，发射机发送的低速率基带信号，经过载波调制后，成为窄带 BPSK 信号，设基带码速率为  $R_b$ ，则窄带信号的第一零点带宽为  $2R_b$ 。再经过扩谱调制后，扩谱信号的带宽主要由高速率的扩谱伪随机序列的速率决定。图 6-4 中给出的是 BPSK 调制，如果伪随机序列的速率为  $R_c$ ，则直扩信号的第一零点带宽则为  $2R_c$ 。一般有  $R_c \gg R_b$ ，因此，扩谱调制后信号的频谱被大大展宽了。

在接收端，直扩信号经过解扩处理后，伪随机序列调制成分被去掉，信号恢复成为一个普通的 BPSK 信号，其带宽为  $2R_b$ 。该信号经过解调后，就可恢复出发送的信息。在接收机中，我们假设接收端已经完成了伪随机序列的同步和载波同步。

图 6-5 给出了解扩处理前后信号功率谱的示意图。假设接收机接收到的信号中除了有用信号外，还包含窄带干扰、白噪声和其他宽带干扰，我们不难看出：接收端的解扩器在组成上，与发送端的扩谱调制器基本相同，只是在输出端多了一个窄带滤波器。因此，对于与接收机产生的伪随机序列同步的有用信号，解扩器将它恢复成为窄带信号，且系统设计使得该窄带信号恰好通过后面的窄带滤波器。由于本地伪随机序列信号是单位功率双极性信号，因此在解扩器中乘法器前后信号的功率是不变的，但是有用信号的带宽被大大压缩，因此其功率谱密度大大提升。对于进入接收机的窄带干扰，解扩器所起的作用是扩谱调制的作用。即窄带干扰被本地伪随机序列调制，成为一个带宽被极大扩展的宽带干扰信号。与有用信号类

似，扩展前后干扰的功率是不变的，因此，在解扩后其功率谱密度大大降低。对于带限白噪声和其他宽带干扰，通过解扩器后，其带宽也同样被扩展，功率谱密度下降，但是其下降的幅度没有窄带干扰那样显著。这样的信号通过后面的带通滤波器后，大部分的干扰功率被滤除，而信号功率基本没有损失，因此解扩后信号的信噪比比解扩前大大提高，实现了抗干扰的功能。

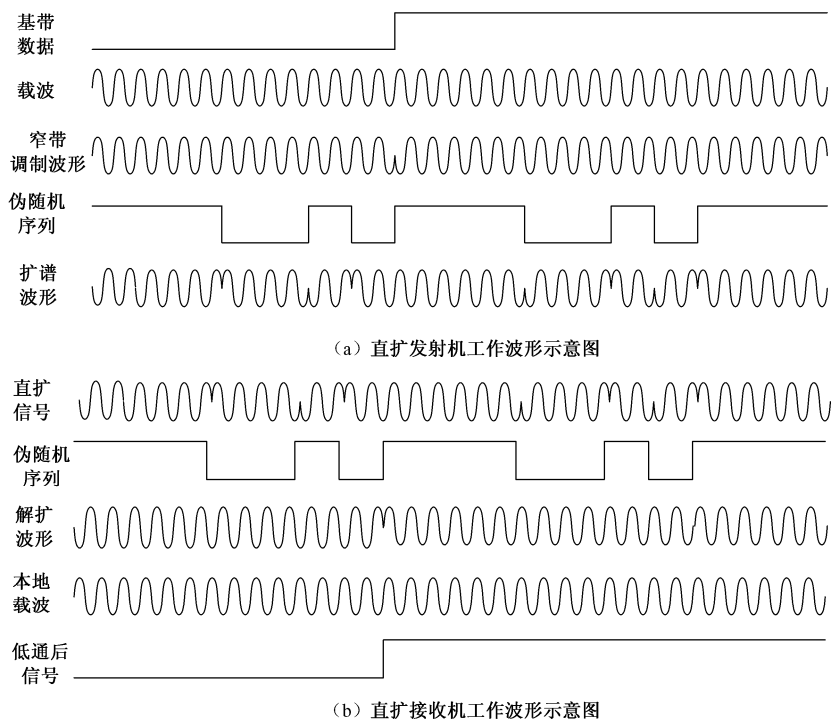


图 6-4 直扩系统的工作波形示意图

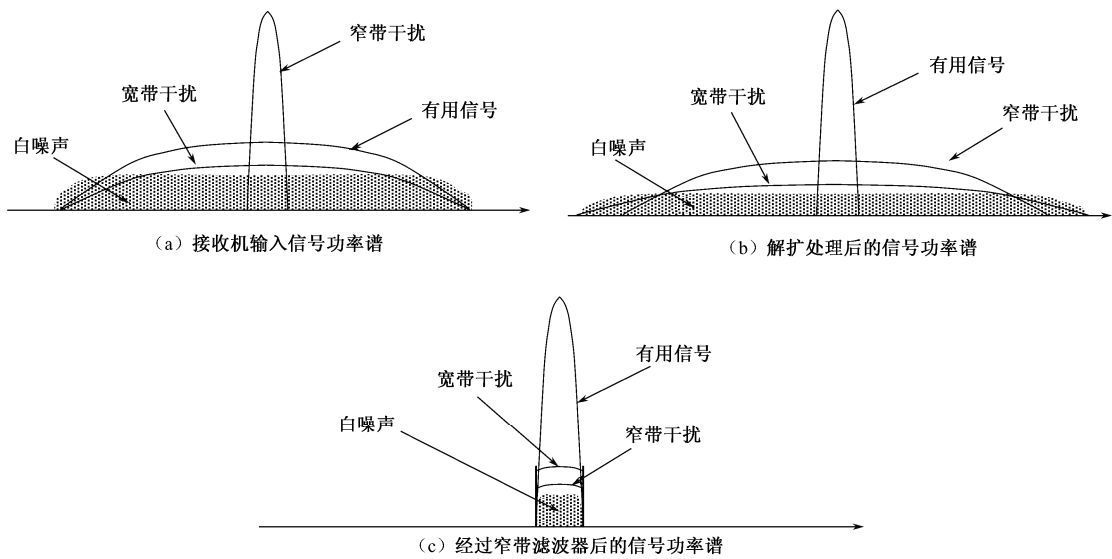


图 6-5 直扩接收机解扩处理前后的信号功率谱变化示意图

2. 跳变频率扩谱（FHSS）

一个普通的窄带通信系统，如果其中心频率在不断变化，那么就是一种跳频通信。实际的跳频通信系统的频率变化是受跳频伪随机序列来控制的，即其频率的变化也遵循着一定的规律。

图 6-6 给出了跳频通信系统的组成方框图。

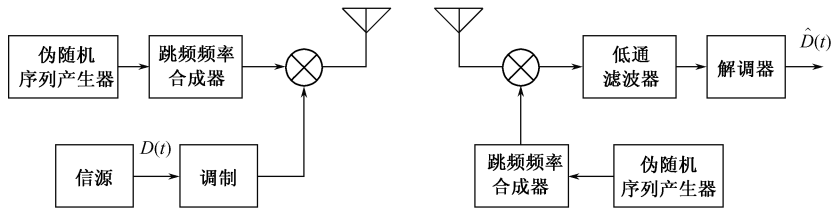


图 6-6 跳频通信系统原理方框图

发送端用伪随机序列控制频率合成器的输出频率，经过混频后，信号的中心频率就按照跳频频率合成器的频率变化规律变化。在接收端的跳频频率合成器与发送端按照同样的规律跳变，这样在任何一个时刻，接收端频率合成器输出的频率与接收信号正好相差一个中频。这样，混频后就输出一个稳定的窄带中频信号。此中频信号经过窄带解调后就可以恢复出发送的数据。与直扩系统一样，跳频系统同样需要同步。

跳频系统在每个频率上的驻留时间的倒数称为跳频速率。当系统跳频速率大于信息符号速率时，称为快跳系统。此时系统在多个频率上依次传送相同的信息，信号的瞬时带宽往往由跳频速率决定。当系统跳频速率小于信息符号速率时，称为慢跳系统。此时系统在每跳时间内传送若干波特的信息，信号的瞬时带宽由信息速率和调制方式决定。目前绝大多数的跳频系统都是慢跳系统。

跳频系统的频率随时间变化的规律称为跳频图案。为了直观地显示跳频系统的跳频规律，可以用图形方式将跳频图案显示出来。图 6-7 给出了一种跳频图案。

该跳频图案中共有 8 个频率点，频率跳变的次序为  $f_3$ 、 $f_1$ 、 $f_5$ 、 $f_7$ 、 $f_4$ 、 $f_8$ 、 $f_2$ 、 $f_6$ 。实际应用中，跳频图案中频率的点数从几十个到数千个，一般认为跳频系统的处理增益就等于跳频点数，如当跳频频率点为 200 个时，其处理增益即为 23 dB。而跳频系统完成一次完整跳频过程的时间也很长，在每个跳变周期中，一个频率有可能出现多次。跳频图案中两个相邻频率的最小频率差称为最小频率间隔。跳频系统的当前工作频率和下一时刻工作频率之间的频差的最小值称为最小跳频间隔。实际的最小跳频间隔都大于最小频率间隔，以避免连续几个跳频时刻都受到干扰。

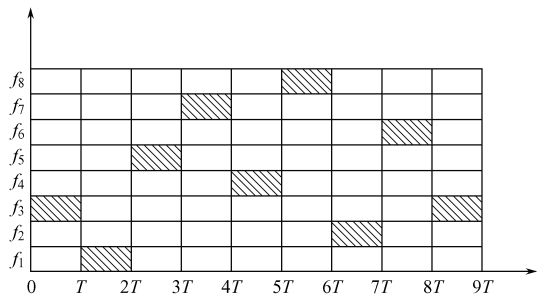


图 6-7 跳频图案示意图



FHSS 的核心是跳频频率合成器，频合可输出的频率点数和跳频转换速率是频合的主要指标，也是决定整个跳频系统性能的主要参数。跳频点数越多则扩展的频谱越宽，系统的处理增益  $G_p$  就越大；跳频速率越高，就可适应高速数据传输，并能更有效地抑制干扰，特别是人为转发式干扰。

由于跳频通信系统在多个频率上跳变，如何理解其干扰容限及其工程意义，有一些问题需要讨论。

根据式 (6-3) 和式 (6-4)，跳频处理增益在理论上表明了跳频通信系统在射频可用带宽内的可用频率个数，也是敌方为了阻塞干扰跳频通信并实现与相同功率干扰常规规定频通信同样效果所要付出的理论代价，因为干扰机的功率需要在每个跳频频率上平均分配。可见，跳频通信理论上的抗阻塞干扰能力就是可用频率数，亦即：在理论上，如果被有效干扰频率数不大于可用频率数，则跳频通信系统应该能正常工作。但是，由于数据压缩、跳频同步和解调过程中的一些实际误差以及器件的非理想性等因素，造成了跳频通信系统的固有损耗，加上跳频通信接收机要维持最小的输出信噪比和工程中跳频同步的维持需要有一定数量频率的支持，使得跳频通信系统的实际抗阻塞干扰能力要小于可用频率数。需要注意的是，在实际跳频通信工程中，尽管每个跳变频率的接收灵敏度可能不尽相同，但实现跳频同步后，各频率均在同一个中频上进行解调，所以所述跳频最小信噪比即为中频解调所需的最小信噪比。例如，按照跳频接收机的一般技术水平，设跳频频率数为  $N$ ，跳频处理固有损耗为 2.5 dB，跳频所需最小解调输出信噪比（或信干比）的门限值为 2.4 dB，代入式 (6-7) 并求反对数，可得跳频干扰容限为 0.32N（百分比）。这就是常规跳频的干扰容限（即实际抗阻塞干扰能力）为可用频率数的 30 % ~ 40 % 的基本原因。当然，该值不是理论值，它是随着技术水平和技术方案的变化而变化的。

实际上，以上范例是以跳频加纠错编码为前提的，若没有纠错编码，跳频干扰容限还难以达到 (30 % ~ 40 %)  $N$ 。因为跳频干扰容限与纠错编码关系十分密切，主要表现在纠错编码可以降低跳频通信系统解调器所需的最小输出信噪比的要求，从而提高跳频干扰容限，或者说无纠错编码跳频通信系统的干扰容限是很低的，亦即跳频本身没有纠错能力，必须与纠错编码同时使用。

总之，跳频通信系统的干扰容限，表现为在可用射频带宽内接收机实际能承受的被有效干扰的频率数，这种表现形式与直扩通信系统的干扰容限是不同的。值得指出两点，一是在跳频干扰中，对于被干扰频率，只要干信比达到了有效干扰的要求，再加大干扰功率没有意义，即跳频通信系统对窄带或部分频带干扰功率的增加不敏感；二是虽然跳频干扰容限及处理增益主要是针对阻塞干扰而言的，与跟踪干扰没有直接的关系，但实际中较大的跳频干扰容限和处理增益会间接地提高跳频通信系统的抗跟踪干扰性能。

### 3. 跳变时间扩谱 (THSS)

跳时扩谱系统即用伪随机序列来键控发射机：即将一个信息码元的持续时间分成若干个时隙，由伪随机序列控制在哪一个时隙中发射信号。由于信号不是连续发射的，跳时扩谱信号的占空比小于 1，一般在伪随机序列控制下，占空比可以达到 50 %。图 6-8 给出了采用 15 位 m 序列 “000100110101111” 控制的跳时信号，其中在码元为 “1” 时开启发射机，在码元为 “0” 时关闭发射机。

由于跳时信号的占空比小于 1，在发射瞬时功率不变的情况下，比特能量降低，对发射机的利用效率降低。因此，跳时技术通常在扩谱系统中不单独使用，而与其他扩谱方式结合组成混合式扩谱方式，如与 FHSS 混合使用，或者利用伪随机序列控制的跳时来实现时分多

址 (Time Division Multiple Access, TDMA)。

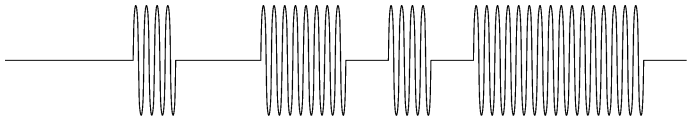


图 6-8 跳时信号波形图

跳时可用于减少时分复用系统之间的干扰，但整个系统有严格的定时要求，以保证发射机之间的重叠最小。而且，和其他扩频通信系统一样，必须选择互相关特性小的伪码序列。跳时系统的处理增益，按定义应该等于信码被划分成的时隙数目或等于占空比的倒数。

#### 4. 调频扩谱 (Chirp SS)

调频信号是一种信号瞬时频率随时间变化的信号，根据频率的递增或递减关系，又分为正斜率和负斜率调频信号两种。图 6-9 给出了正斜率和负斜率两种调频信号。当频率的递增或递减与时间成线性关系时，称为调频扩谱信号，又被称为 Chirp 信号。

调频扩谱系统是基于调频信号产生和压缩的扩谱系统，以及基于模拟信号调频扩谱，由于调频信号在压缩过程中对多普勒频移不敏感，因此被广泛应用在脉冲压缩体制的雷达系统中。



图 6-9 调频扩谱信号

调频扩谱信号可以表示为

$$s(t) = \text{rect}\left(\frac{t}{T}\right) \cos\left(\omega_0 t + \frac{1}{2} \mu t^2\right) \quad (6-10)$$

式中， $\text{rect}\left(\frac{t}{T}\right) = \begin{cases} 1 & \left|\frac{t}{T}\right| \leq \frac{1}{2} \\ 0 & \left|\frac{t}{T}\right| > \frac{1}{2} \end{cases}$  为宽度为 1 的矩形函数； $\omega_0$  为中心频率； $T$  为调频扩谱信号的

时宽； $\mu$  为调频斜率， $\mu = \pm 2\pi B/T$ ，其中的正负号对应调频扩谱信号的正负斜率； $B$  为调频扩谱信号的调频带宽，当信号的时宽带宽积很大时， $B$  约等于信号的实际带宽。

由式 (6-10)，调频扩谱信号的频谱为

$$\begin{aligned} S(\omega) &= \int_{-\infty}^{\infty} s(t) \exp(-j\omega t) dt \\ &= \frac{1}{2} \int_{-\frac{T}{2}}^{\frac{T}{2}} \exp\left\{j\left[(\omega_0 - \omega)t + \frac{1}{2} \mu t^2\right]\right\} dt \\ &\quad + \frac{1}{2} \int_{-\frac{T}{2}}^{\frac{T}{2}} \exp\left\{-j\left[(\omega_0 + \omega)t + \frac{1}{2} \mu t^2\right]\right\} dt \end{aligned} \quad (6-11)$$

从式 (6-11) 可以看出, 信号的频谱集中在  $\pm\omega_0$  附近。通常情况满足  $\omega_0 \gg 2\pi B$ , 可以认为正负两部分频谱互不重叠。

由上述讨论的结果可见, 调频扩谱信号的格式比较简单, 但是同样可以达到类似于 DS 信号的结果—通过匹配解扩提高信号的信噪比。一般调频扩谱系统的处理增益认为等于调频信号的时带积, 这刚好对应雷达系统中常用的概念——压缩比。调频扩谱信号经过匹配滤波输出的连续相关峰信号的处理, 与其他扩谱信号匹配后处理的过程类似。

### 5. 混合扩展频谱 (Hybrid SS)

前讨论了四种基本扩频系统, 它们各有优缺点。在系统设计中若仅用一种本调制方式, 往往达不到使用性能上的要求, 若将两种或多种基本的扩展频谱方法结合起来, 结合各自的优点, 就能得到只使用一种扩频方法所不能达到的性能, 甚至有可能降低系统的实现难度。常用的混合扩展频谱调制方式有: 跳频和直扩的混合调制 (FH/DS)、调频扩谱和直扩的混合调制 (Chirp/DS)、跳时和跳频的混合调制 (TH/FH)、跳时和直扩的混合调制 (TH/DS)。这里仅对前两种混合扩谱系统进行介绍。

#### 1) 跳频/直接序列混合扩频系统 (FH/DS)

在 FH/DS 系统中, 扩展频谱信号是由载波频率跳变的直接序列信号组成。图 6-10 给出了系统的原理方框图, 图 6-11 给出了信号的频谱图。

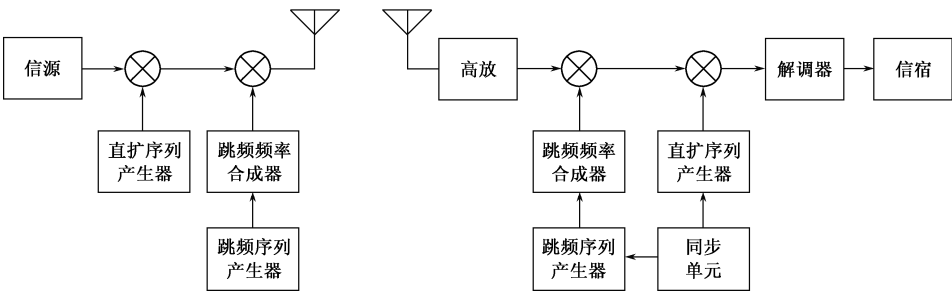


图 6-10 FH/DS 混合扩谱系统原理方框图

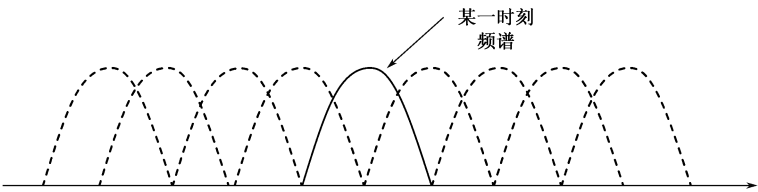


图 6-11 FH/DS 信号频谱图

由图 6-10 可见, FH/DS 系统的发射机是直接序列调制器与跳频载波的直接叠加, 这种混合调制器与单纯的直接序列调制的区别在于, 它的载波是按跳频图案变化的, 而不像单纯 DS 调制那样是固定不变的; 与单纯的跳频系统的区别在于, 每次跳变必须把直扩 ( $\sin x/x$ ) 型功率谱依跳频图案的规则搬移, 且跳变频道的最小间隔必须等于直接序列的伪码速率 (目的在于保证相邻跳变频道之间的正交性)。图 6-10 中的跳频伪随机序列发生器向频率合成器提供跳频图案, 直扩伪随机序列产生器向平衡调制器提供 DS 扩谱调制序列, 故平衡调制的输

出就是 FH/DS 信号。FH/DS 系统在任何一个跳频时间内都是一个直扩系统，即其瞬时频谱也是宽带的，而在一个跳频周期内它是由许多直接序列扩频信号的频谱构成的。占有一定带宽的直接序列信号按照跳频图案（即时频矩阵图）伪随机地出现，每个直接序列信号在发射瞬间只覆盖系统总带宽的一部分频段，采用这种混合系统能够提高抗干扰能力并简化设备，降低分机部件技术难度，并能达到多址的目的。

FH/DS 系统的处理增益为跳频和直扩系统的处理增益的乘积，如果用分贝表示则是相加的关系，即

$$G_{\text{FH/DS}} = G_{\text{FH}} + G_{\text{DS}} \tag{6-12}$$

式 (6-12) 中， $G_{\text{FH}} = N$  为 FH/DS 中的可用跳频数； $G_{\text{DS}} = R_c / R_b$ 。

当单纯使用跳频或直扩难以达到指定的扩谱带宽：即单纯用直扩则伪码发生器时钟速率  $R_c$  已达到最大值，或者跳频器的可用跳频数已达到最大限度，那么使用混合式 FH/DS 调制特别有价值。例如，某系统要求扩展频谱后的射频带宽为 1 000 MHz，数据速率为 5 kb/s。若采用 DSSS，则要求伪码发生器的工作速率为 500 Mb/s；若采用 FHSS，则要求跳频器输出的跳频数为  $N=100\,000$ ，跳频间隔为 5 kHz。这样高的伪码速率或这么大的跳频数，在技术实现上困难极大。但是，如果采用 FH/DS 系统，则用 50 Mb/s 的伪码发生器和一个只需跳频数为 20、最小频率间隔为 50 MHz 的频率合成器就能满足系统要求。显然，这种混合系统的各部件技术难度就降低了，便于工程实现。

使用 FH/DS 混合扩频的主要原因还在于它综合了 FH 和 DS 的优点，消除了它们单独使用的弱点。如 DSSS 存在抗单频瞄准式干扰和远近干扰能力差的缺点；FHSS 抗宽带干扰和中继转发式干扰差；而 DS/FH 混合扩频调制在抗单频干扰、宽带干扰及中继干扰好。

## 2) 直扩/调频扩谱混合扩谱系统 (DS/Chirp)

直扩信号在每个伪随机序列码元周期内都是一串连续的正弦波，而 DS/Chirp 混合扩谱信号在每个伪随机序列的码元周期内都是一个调频扩谱信号。如图 6-12 所示，可见在每个码元周期内的调频扩谱信号的相位受伪随机序列的控制，形成 DS/Chirp 混合扩谱信号。

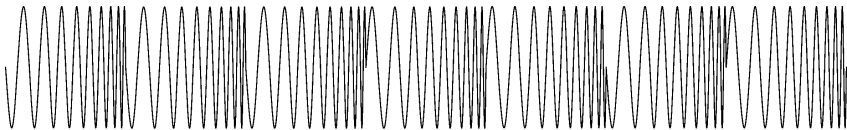


图 6-12 DS/Chirp 混合扩谱信号

图 6-13 是一种 DS/Chirp 混合扩谱系统的原理方框图。在发送端数据经过差分编码后，再对 DS/Chirp 混合扩谱信号调制。在接收端的处理与调频扩谱系统类似，用匹配滤波器对混合扩谱信号进行解扩，输出的相关峰信号的相位变化用差分相干解调。

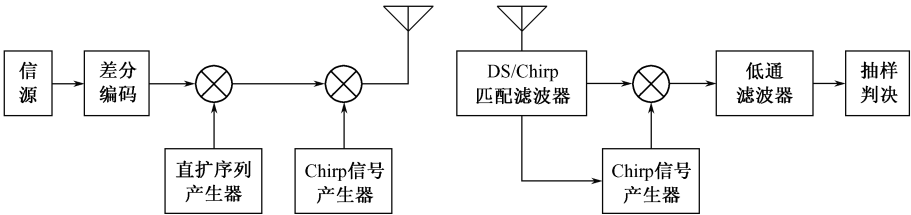


图 6-13 DS/Chirp 混合扩谱系统原理方框图

3) 跳时/直接序列 (TH/DS) 混合扩频系统

当 DSSS 中使用的独立地址码数, 不能满足多址和复用的要求时, 在 DSSS 中增加时分复用 (TDM) 是一种有效的办法。由于 DSSS 中收、发两端之间已有准确的伪码时钟同步, 它已经具有很好的定时, 足以维持时分复用正常工作, 这就为增加 TH 技术带来了方便之处。因此, 在 DSSS 中增加 TH 工作时, 只需要增加一个通断开关及有关的控制电路即可。图 6-14 绘出了 TH/DS 系统的原理方框图。

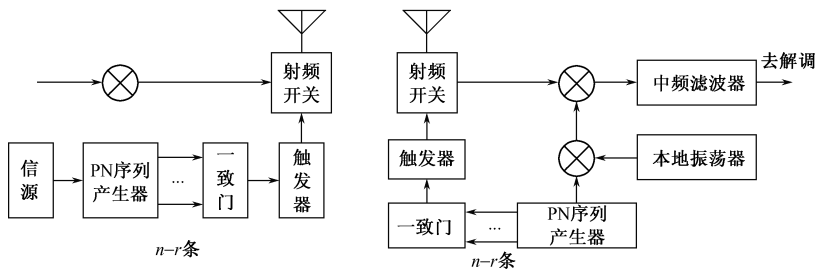


图 6-14 TH/DS 系统的原理方框图

对于跳时工作来说, 启闭判可以很容易地从直接序列扩频码发生器得到。发射机的  $n$  级伪码发生器输出除供给直接序列作载波平衡调制外, 还从  $n$  级伪码发生器中另选  $(n-r)$  级的状态并行输出到一致门 (又称符合门)。当它们都处于 “1” 状态时, 即控制射频开关发出脉冲载波信号。在伪码一个周期中,  $(n-r)$  级出现全 “1” 的状态为  $2r$  次, 也就是说发射机在一个伪码周期中发射  $2r$  次 (这里  $1 < r < n$ ), 而且全 “1” 状态出现的分布是伪随机的, 因此发射也是伪随机的。

接收机工作状态与发射机类似, 只要用与所接收的发射信号同步的控制信号去启闭接收机的前级使其正常工作即可。方框图中未画出 DSSS 中所需的同步跟踪环路, 另外接收机启闭时间也由环路来控制伪码发生器, 解决传播延迟的影响。

TH/DS 系统中由于在 DSSS 的基础上加入了跳时, 这一伪随机的时间分配可以容纳更多信道的用户, 与单纯的码分复用相比, 改善了工作状态, 增加了独立地址数, 对近台干扰抑制的效果也好。

6.3 通信干扰技术

6.3.1 通信干扰技术的分类

通信干扰按不同方法可以有多种分类, 经常用到的分类方法主要有以下几种。

1. 按作用性质分

通信干扰按作用性质可分为欺骗性干扰和压制性干扰。

欺骗性干扰又称迷惑性干扰, 它是模拟敌方的通信信号来欺骗敌方, 使其作出错误的判断和决策。实现欺骗干扰时, 常常模仿敌人的一个外站, 进入敌方的通信网络, 根据推测发送坏的信息或错误的命令, 也可模仿敌方更高一级指挥员的声音对敌军下达命令。还可以在

敌通信链路上发送混乱的信息和声音，使敌无线电台操作人员的工作效率降低，比如，女人的声音，单音背诵声，流行音乐和灾难报告，等等。

一般来说，欺骗性干扰较难获得成功，它要求干扰信号与敌通信信号要极其相似，需要充分掌握敌通信电台的技术和战术特点、通联规律等资料。为了获得更好的效果，有时还需要其他兵种的协同。由上可见，欺骗性干扰的有效性主要取决于战术上的运用。对于采用密码技术的通信系统，实现欺骗性干扰比较困难。从技术的角度看欺骗性干扰与通信本身没有太大的区别。欺骗干扰战术上的运用可参考其他有关战术方面的资料，本书将着重讨论压制性干扰的技术问题。

压制性干扰就是人为地发射干扰电磁波，使敌方的通信接收设备难于或完全不能正常接收通信信息。它是以强的干扰掩盖通信信号，致使通信接收机降低或丧失正常接收信息的能力。有效的压制性干扰将使敌方接收机接收到的信号模糊不清或完全被掩盖，它是一种强有力的人为积极干扰，是通信干扰研究的主要对象。

2. 按同时干扰信道数分类

通信干扰按同时干扰信道的数目可分为拦阻式干扰和瞄准式干扰。

拦阻式干扰又称阻塞式干扰，是同时对某个频段内多个或全部信道的干扰，干扰的作用带宽等于目标信号的工作频率范围，或者覆盖目标信号的部分工作频率范围。由于干扰功率扩展在其覆盖的所有信道中，这种干扰技术通常要求干扰机具有大的输出功率或近的干扰距离，以保证在每个信道中的干扰功率足以压制通信信号。

我们知道通信干扰应采用同频干扰，只有将干扰的频率重合到信号的频率上才能形成同频干扰，瞄准式干扰正是这样一种干扰技术，它是针对一个无线电信道的同期干扰。与拦阻式干扰相比，瞄准式干扰的功率利用率更高，但干扰方需要掌握目标信号的中心频率及带宽的信息，使干扰信号的中心频率及频谱宽度与目标信号的中心频率及带宽相重合才能达到有效的干扰。

根据干扰瞄准信号程度的不同，瞄准式干扰又分为准确瞄准式干扰和半瞄准式干扰。通常把干扰与目标信号频谱重合程度高于 75 % 的干扰称为准确瞄准式干扰，也称为瞄准式干扰，如图 6-15 所示。而把干扰与信号频谱不重合或不完全重合，但 75 % 以上的干扰能量能通过接收机的选择电路的干扰称为半瞄准式干扰，如图 6-16 所示。一般准确瞄准式干扰比半瞄准式干扰的干扰效果好，但其对频率重合度的要求较高，设备复杂。在某些情况下如对快连通信信号干扰时实现准确瞄准式干扰比较困难。但无论准确瞄准式干扰或半瞄准式干扰，它们的工作原理、目的是相同的，所不同的仅仅是干扰与信号频率的重合度不同。在设备和条件允许的情况下，我们总是希望干扰与信号的频率重合度越高越好。

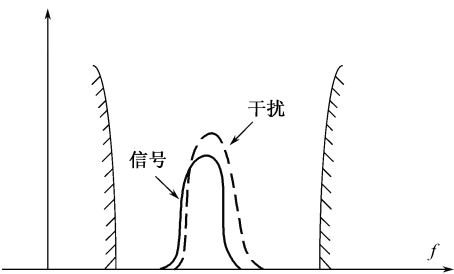


图 6-15 瞄准式干扰示意图

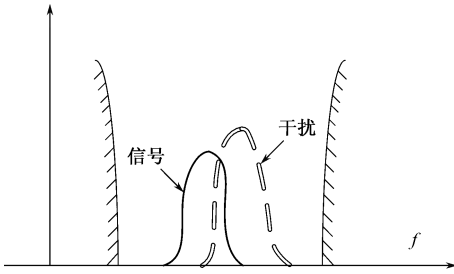


图 6-16 半瞄准式干扰示意图

在瞄准式干扰中按控制与实施干扰方式的不同又可分为转发式干扰、跟踪式干扰、扫频搜索式干扰、间断式干扰、连续式干扰及单一频率瞄准干扰、信号电平启动触发式干扰等。随着智能化控制技术的发展,这种分类有越分越细,越分越多的趋势,但同时一部干扰机所具有的干扰方式也越来越多。

### 3. 按干扰平台分类

按干扰机所在的平台分类,有便携式、车载式、机载式、艇载式、摆放式、投掷式干扰机等。

陆军通信干扰机大都采用车载式干扰机,它具有干扰功率较大,机动性强等优点。一些小型干扰机则可采用战斗员背负便携式。就发射信号来说,通信干扰机与通信发射机没有本质的区别。除了专用的干扰机外,还可使用干扰附加器,利用通用的无线电发射设备进行通信干扰。

同其他人为积极干扰一样,通信干扰机一旦开始工作就暴露了自己而容易遭到敌方的攻击。另一方面,为了提高干扰功率总希望干扰机离目标接收机越近越好,因此总是将干扰机尽可能地部署到阵地前沿。所以,经常采用遥控的或预编程的远距离干扰机,以有利于保护指挥部及干扰操作员的安全。

无人干扰飞机能在靠近目标的近距离点上实施干扰,而不会对操作员产生危险。它能提供到作用目标较佳的视距干扰而提高其干扰效率。如果无人干扰机飞的足够高,地形对电波传播的影响就非常小,在同等干扰功率的情况下,无人干扰飞机比地面干扰机更有效。飞机或者任何有人驾驶的空中载体也可运载干扰机,称其为机载干扰机,具有与无人干扰飞机同样的性能。一般机载干扰机被当作合成部队协同作战的力量来部署,这些空中资源的部署要受到上一级指挥部的控制。一般来说,当空军使用其资源来支援电子战作战任务时,应该通知地面部队可能会在哪个具体时间、具体地点出现干扰。

人工摆放式干扰机或者一次性使用干扰机,被秘密地放置在靠近目标的地方或者摆放在阵地前沿,在关键时刻遥控触发其干扰敌人的通信。其触发工作方式有有线方式、无线方式和定时器触发方式几种。由于这时干扰机非常靠近目标,所以可大大提高干扰效率。但是这种干扰机的天线高度一般都很低,所以又抵消了一部分好处。

投掷式干扰机是一种一次性使用的干扰机,这种干扰机简单而坚固,一般用火炮投掷到敌方地域内。它们是定时工作的,寿命很短,一旦其电源用完,应能自毁。

此外,通信干扰还有许多其他的分类方法:按电波传播方式,通信干扰可分为地波干扰、天波干扰和空间波干扰,按干扰机的工作频段可分为长波、中波、短波、超短波和微波干扰,按干扰作用时间可分为连续式干扰和间断式干扰,按设备使用分干扰附加器、专用干扰机、投放式干扰机、一次性使用干扰机,等等。除此之外,通信干扰还有按干扰强度、干扰信号形式、调制方式、作用距离等分类。这些分类的意义都一目了然,不再详述。

## 6.3.2 通信干扰的理论基础

### 1. 干信比

在通信对抗中,通信干扰机的效果只有与受干扰的目标接收机关联起来方可进行评估。

干扰效果常用有效干扰功率（即进入目标通信接收机核心部件的干扰信号功率）与信号功率（接收机所接收到的有用信号功率）的比值来描述，这一比值被称为干扰信号比或干信比，简单表示为  $J/S$ ，通常用 dB 来表示。

为进一步说明干信比的基本概念和应用，这里将首先给出通信干扰的位置分布示意图，如图 6-17 所示。和雷达不同，由于通信链路的作用就是将信息从一个地点传递到另一个地点，所以，一般来说，通信接收机总是位于和发射机不同的地点。

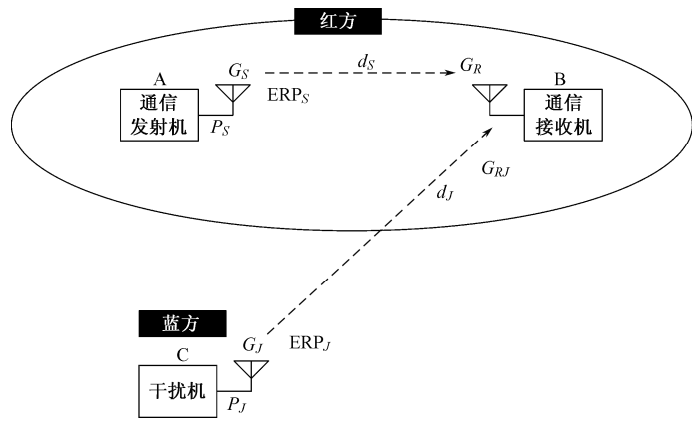


图 6-17 通信干扰位置分布图

干扰机干扰通信的途径是向目标接收机注入不希望接收到的信号，该信号利用无线信道的开放传输特性，伴随着通信信号一起进入接收机。为达到干扰目的，这个不希望接收到的信号必须足够强，使得接收机无法从接收的混合信号中恢复出所需信息。下面，我们就根据图 6-17 来给出  $J/S$  的表达式。

首先考虑有效的通信信号部分。在信号从发射机单程传输到接收机的情况下，如图 6-17 中的 A 和 B 之间的链路关系。根据链路传播损耗公式，我们可以得到到达接收机输入端的信号功率为

$$S = P_S + G_S - L_S + G_R \tag{6-13}$$

其中， $P_S$  为通信发射机的发送功率，单位为 dBm； $G_S$  为通信发射机的发射天线增益，单位为 dB； $G_R$  为通信接收机的接收天线增益，单位为 dB；而  $L_S$  为发射机与接收机之间的路径损耗，单位为 dB，满足

$$L_S = 32.4 + 20 \lg d_S + 20 \lg f \tag{6-14}$$

其中， $d_S$  为收发信机之间的距离，单位为 km； $f$  为发射信号的频率，单位为 MHz。

下面我们来看一看到达通信接收机的干扰信号部分。从图 6-17 不难看出，干扰信号实际上是单程传输的。一般来说，无论干扰目标是通信接收机还是雷达接收机，干扰信号的性能都是相同的。接收机对干扰信号的接收不同于在双程传输情况下对所需信号的接收情况。首先，除非接收机采用全向天线，否则天线增益将随接收信号的方位和仰角变化而变化。因此，如果不是从同一个方向到达，干扰信号和通信信号将遇到不同的接收天线增益。其次，因为无法测量或估计通信信号的确切频率，所以干扰信号的频带宽度就必须远大于要干扰信号的带宽。预测  $J/S$  时，只考虑落入接收机工作带宽之内的干扰信号。根据上述两点考虑，到达接收机输入端的干扰功率可以由下式给出：



$$J = P_J + G_J - L_J + G_{RJ} \quad (6-15)$$

其中,  $P_J$  为干扰机在接收机带宽内的发送功率, 单位为 dBm;  $G_J$  为干扰机的发射天线增益, 单位为 dB;  $G_{RJ}$  为通信接收机在干扰信号方向的接收天线增益, 单位为 dB; 而  $L_J$  为干扰机与接收机之间的路径损耗, 单位为 dB, 同样满足

$$L_J = 32.4 + 20\lg d_J + 20\lg f \quad (6-16)$$

由于  $J$  和  $S$  是以 dB 表示的, 所以二者的比值其实就是它们之间的 dB 值之差。因此, 对于通信干扰来说,  $J/S$  可以表示为

$$\begin{aligned} J/S &= (P_J + G_J - L_J + G_{RJ}) - (P_S + G_S - L_S + G_R) \\ &= P_J - P_S + G_J - G_S - 20\lg d_J + 20\lg d_S + G_{RJ} - G_R \end{aligned} \quad (6-17)$$

例如, 若蓝方干扰机的反射功率为 100 W (50 dBm), 其天线增益为 10 dB, 距红方接收机的距离为 30 km; 红方发射机与其接收机的距离 10 km, 发射功率为 1 W (30 dBm), 发射天线增益为 3 dB, 接收机天线对通信信号和干扰信号的增益均为 3dB。那么根据式 (6-17), 我们很容易求出  $J/S$  为

$$\begin{aligned} J/S &= 50 - 30 + 10 - 3 - 20\lg 30 + 20\lg 10 + 3 - 3 \\ &= 27 - 20\lg 3 \approx 17.5 \text{ (dB)} \end{aligned} \quad (6-18)$$

## 2. 通信干扰方程

一旦通信干扰机和被干扰通信系统的位置确定以后, 在干扰实施过程中, 干扰方总希望能够获得干扰目标处的干信比, 以判断干扰的有效性。但在实际工作中, 对于干扰方来说是很难直接确定目标接收机的位置的。一般来说, 在实际的通信系统中, 某一通信台站同时都要进行收、发工作, 尤其是对战术通信电台更是如此, 这样就可以借助电子支援措施(ESM)来获得被干扰接收机的位置信息。对战略通信电台而言, 虽然可以利用遥控使同一端的发射机与接收机不在同一地点, 但收发信机的距离通常比干扰距离小得多, 可将发射机的位置大致定为接收机的位置。本节的目的就是要给出干扰机与其作用目标通信链路之间相互关系的方程, 从而确定目标接收机处的干信比, 以及进行干扰功率和作用距离的估算。

假设目标接收机的带宽是理想的, 且该接收机调谐到通信信号频率上, 那么到达接收机的信号功率仅存在传输损耗, 如式 (6-13)、(6-14) 所示。而干扰功率除了传输损耗以外, 还可能存在其他损耗。这些额外损耗主要来自两方面的影响:

一是滤波损耗, 这是由接收机的带通滤波器引起的。当干扰信号带宽大于接收机带宽或者干扰频率偏离信号频率时, 都会降低有效干扰功率, 因为接收机将抑制通带以外的干扰, 使一部分干扰功率浪费掉了。滤波损耗  $F_b$  的对数形式定义为

$$F_b = 10\lg \left( \frac{\text{进入接收机的干扰频谱宽度}}{\text{干扰频谱带宽}} \right) \quad (6-19)$$

滤波损耗反映了进入目标接收机的有效干扰功率占干扰总功率的比例, 当全部干扰功率都进入接收机时, 滤波损耗  $F_b = 0 \text{ dB}$ 。

二是可能存在的极化损耗, 这是由于干扰机可能不是以合适的极化电波发射干扰信号造成的, 可以用系数  $\rho$  来表示,  $\rho$  的取值范围为 0~1, 换算为对数形式, 则不超过 0 dB。

考虑滤波损耗和极化损耗以后, 接收机处的干扰功率  $J$  (以对数表示) 可以修正为:

$$J = P_J + G_J - L_J + G_{RJ} + F_b + \rho \quad (6-20)$$

我们再来结合图 6-19 所示的干扰机、通信反射机和通信接收机的地理位置关系，讨论目标接收机处的  $J/S$ 。对于瞄准式干扰而言，干扰信号带宽一般近似等于目标信号带宽，因此，可取  $F_b = 0$ 。极化损耗的影响难以准确估计，通常是在计算干扰功率时留有适当的功率余量加以弥补，于是我们就可以得到：当目标接收机处的干扰功率不小于  $k$  倍信号功率时，干扰功率必须满足如下公式：

$$P_J \geq K + P_S + G_S + G_R + 20\lg d_J - G_J - 20\lg d_S - G_{R_I} \quad (6-21)$$

其中， $K = 10\lg k$ ，单位为 dB。

式 (6-21) 就被称为干扰方程。其中， $K$  和  $G_J$  是由干扰方确定的已知量，通信方的发射功率和天线增益只能根据 ESMT 提供的情报来进行估算。干扰与通信的路径损耗也只能粗略概算。由此可知，利用干扰方程只能粗略概算所需要的干扰功率。这种概算对于确定干扰功率或合理配置干扰站都是具有实用意义的。

### 3. 最佳干扰概念

从上面分析可以看出， $J/S$  不仅与干扰功率有关，它还和许多其他因素有关。因此，对于不同的无线电通信方式来说，就有着不同的最佳干扰样式。为了描述干扰信号的参数与被干扰接收机失真程度的关系，我们引入了最佳干扰的概念。所谓最佳干扰，就是被干扰通信接收机处的  $J/S$  达到一定程度时，使敌接收机接收信号造成最大失真或最大误码率的干扰。也就是说，针对某种信号形式的给定接收方式，能以最小干扰功率达到最好干扰效果的干扰就是对这种信号形式、接收方式的最佳干扰。

衡量最佳干扰的主要参数为压制系数。它是根据目标接收机被有效压制时，被干扰的接收机输入端的最小  $J/S$ ，即

$$K_J = (J/S)_{\min} \quad (6-22)$$

从式 (6-22) 可以看出，压制系数  $K_J$  越小，说明我们可以用较小的干扰功率就能压制敌方的无线电通信。压制系数的大小与下列条件有关：

- 频率重合度（干扰频率和信号频率重合的程度，对瞄准式干扰而言）；
- 目标通信信号的结构和接收机的接收方法以及接收机性能；
- 压制敌方通信的标准（如误码率要求）。

上述前两个条件不难理解，而第三个条件指的是干扰敌方通信造成的误码率要求，这要根据不同的通信样式和作战要求来确定。

### 6.3.3 对常规通信信号的干扰

在对通信干扰有了初步的认识以后，本节我们对常规通信信号的干扰进行分析研究。为了与通信信号相对照，我们按调制方式划分干扰样式，并找出对主要常规通信信号的最佳干扰样式。

与采用信噪比来衡量通信的质量一样，常用干信比作为衡量干扰对信号影响的程度。由于无线电发射机的功率受限于其峰值功率，有时也采用干扰峰值功率与信号的峰值功率之比，称为峰值干信功率比。一般衡量干扰效果时，应将峰值干信功率比转换为平均干信功率比。在干扰实验中，由于受到测量仪器的限制，有时也采用干扰电压与信号电压之比来衡量干扰效果，这时称其为干信电压比。在以后的讨论中，除特殊说明外，书中的干信比均指平均干

信功率比。

对模拟通信信号和数字通信信号的干扰分析分别采用不同的方法。对模拟通信的干扰分析，是在一定的输入干信比下，用接收机输出的干信比大小作为衡量干扰效果的标志。而干扰数字通信时，则是在一定的输入干信比下，用产生的误码率来衡量干扰的效果。

本节将首先定性给出对模拟信号和数字信号进行干扰的基本原则，然后分别以模拟调制中的 AM 信号和数字调制中的 2FSK 信号为例，详细讨论不同干扰样式对其的影响情况，至于其他常规通信信号，如模拟调制中的 FM，数字调制中的 2ASK、2PSK、2DPSK 等信号，分析方法类似，感兴趣的读者可以参考相关资料或自己进行分析，这里不再赘述。

1. 常规信号的干扰原则

当干扰模拟调制信号时，通常要达到一个较高的干信比（一般认为 10 dB 比较合适）。通常必须以 100 % 占空比进行干扰。这点是必须的，因为接收机操作员的一个重要能力就是“自适应”地收听。在所有的模拟通信中，即使在低质量的信息传输中，我们也能够根据上下文来进行“填空”，猜测出电文的意思。在战术军事通信中更是如此。

当干扰数字调制信号时，我们攻击信号的目的就是让它无法被数字解调器解调。达到这个目的，既可以干扰数字调制的同步部分，也可以干扰其解调判决部分，造成其误码率升高。一般来说，数字通信系统的同步部分较为可靠，所以我们干扰的根本方法还是直接造成其误码率升高。

图 6-18 给出了一个经典的误码率随接收机信噪比而变化的曲线关系。图 6-18 中信号质量用误比特率与  $E_b/N_0$  的关系来表示。

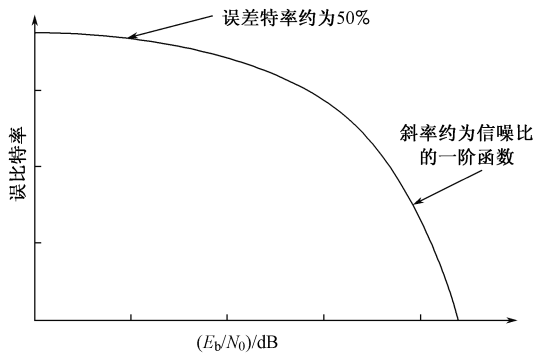


图 6-18 误比特率与信噪比的关系示意图

各种调制的误比特率曲线都类似图 6-6 中这个典型示例的基本形状。曲线顶部在变平，此时的误比特率约为 50 %——恰是数字传输中能达到的最坏状况。所有曲线都在  $E_b/N_0 = 0$  dB 左右信噪比（即信号功率等于噪声功率）时达到了 50 % 这个点。这意味着无论采用那种调制类型，如果噪声电平（或干扰电平）等于接收信号电平，当干扰电平继续增大时误比特率将不再增加很多。

此外，如果信号在三分之一时间是不可读的（短期内），则它被认为是传输无效的。这意味着一个数字信号只需要 0 dB 干信比进行三分之一时间的干扰，基本就可以达到干扰的目的。而对于模拟信号来说，要达到同样目的，需要用正的干信比进行 100 % 时间的干扰。在一些著作中讲到完全干扰数字信号所需干扰占空比是 20 %，但更为普遍认可的是 33 % 这个值。

值得提出的是，数字通信中纠错码的使用大大增加了数字信号的抗干扰能力，由于它们纠正了一些由于干扰引起的比特错误，所以此时所需的干扰占空比就会增加很多。

## 2. 对 AM 信号的干扰

我们知道对 AM 信号的解调可采用非相干解调或相干解调，实际中 AM 解调更多的是利用包络检波器进行非相干解调。下面就对 AM 信号采用非相干解调时的干扰进行分析。

任意 AM 信号可以表示为

$$S_{AM}(t) = [A + m(t)] \cos(\omega_c t + \varphi_s) \quad (6-23)$$

其中， $m(t)$  为基带信号， $A$  为信号载波幅度， $\omega_c$  为载波频率， $\varphi_s$  为信号的初始相位。为分析方便，设  $\varphi_s = 0$ ，则有

$$S_{AM}(t) = [A + m(t)] \cos \omega_c t \quad (6-24)$$

干扰信号一般形式可以写成

$$j(t) = J(t) \cos[\omega_j t + \varphi_j(t)] \quad (6-25)$$

其中， $J(t)$  为干扰振幅， $\omega_j$  为干扰载频， $\varphi_j(t)$  为干扰相位。并假设干扰能够通过接收机的通带而不被抑制。则在接收机中信号与干扰的合成信号为

$$\begin{aligned} x(t) &= S_{AM}(t) + j(t) \\ &= [A + m(t)] \cos \omega_c t + J(t) \cos[\omega_j t + \varphi_j(t)] \\ &= \left\{ A + m(t) + J(t) \cos[(\omega_j - \omega_c)t + \varphi_j(t)] \right\} \cos \omega_c t \\ &\quad - J(t) \sin[(\omega_j - \omega_c)t + \varphi_j(t)] \sin \omega_c t \end{aligned} \quad (6-26)$$

利用三角恒等式可得

$$x(t) = R(t) \cos[\omega_c t + \theta(t)] \quad (6-27)$$

其中

$$R(t) = \left\{ [A + m(t)]^2 + J^2(t) + 2[A + m(t)]J(t) \cos[(\omega_j - \omega_c)t + \varphi_j(t)] \right\}^{1/2}$$

为瞬时幅度，假设理想包络检波器系数为 1，则包络检波器的输出为

$$x_0(t) = [A + m(t)] \left\{ 1 + 2 \frac{J(t)}{A + m(t)} \cos[(\omega_j - \omega_c)t + \varphi_j(t)] + \frac{J^2(t)}{[A + m(t)]^2} \right\}^{1/2} \quad (6-28)$$

以  $A + m(t)$  为参数将此平方根展开成在原点的泰勒级数。当  $A + m(t) \geq 2J(t)$  时，将此级数截短，只取前三项，此时仅有相当小的误差，可得

$$\begin{aligned} x_0(t) &\approx A + m(t) + J(t) \cos[(\omega_j - \omega_c)t + \varphi_j(t)] + \\ &\quad \frac{J^2(t)}{4[A + m(t)]} - \frac{J^2(t)}{4[A + m(t)]} \cos[2(\omega_j - \omega_c)t + 2\varphi_j(t)] \end{aligned} \quad (6-29)$$

式 (6-29) 表明了干扰与信号同时作用于检波器输入端时，在干扰幅度小于信号幅度时的检波器输出情况。由此来讨论不同干扰信号对 AM 信号的干扰。

此时，式 (6-25) 的干扰信号为

$$j(t) = A_{mj} \cos[\omega_j t + \varphi_j] \quad (6-30)$$

代入式 (6-29) 中, 则得检波器的输出为

$$\begin{aligned}
 x_0(t) \approx & A + m(t) + A_{mj} \cos \left[ (\omega_j - \omega_c)t + \varphi_j \right] \\
 & + \frac{A_{mj}^2}{4[A + m(t)]} - \frac{A_{mj}^2}{4[A + mv]} \cos \left[ 2(\omega_j - \omega_c)t + 2\varphi_j \right]
 \end{aligned}
 \tag{6-31}$$

由于平方项很小, 忽略式 (6-31) 中的后两项。此时干扰项为单频干扰, 对于人耳收听的 AM 信号, 不易产生好的干扰效果。当干扰频率与载波频率近似一致时, 检波器输出的干扰项是一缓变的直流分量, 它会被检波器的隔直流电容器所抑制。

## 2. 对数字调制信号 (2FSK) 的干扰

二进制频移键控 (2FSK) 信号是用两个不同频率的载波振荡来传输数字信息 1 和 0。其表达式为

$$u_s(t) = \begin{cases} U_s \cos \omega_{s1}(t), & \text{发 1} \\ U_s \cos \omega_{s0}(t), & \text{发 0} \end{cases}
 \tag{6-32}$$

对 2FSK 信号有不同的解调方法, 我们主要分析对非相干解调 2FSK 信号的干扰情况。非相干解调是 2FSK 信号中最常用的解调方法之一, 其原理框图如图 6-19 所示。其中的两个窄带分路滤波器可分别滤出频率为  $f_{s1}$  和  $f_{s0}$  的载波信号, 包络检波器取出其幅度信息, 输出采用择大判决准则。

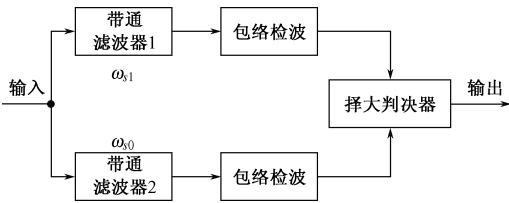


图 6-19 2FSK 非相干解调原理框图

若用 2FSK 干扰信号去干扰 2FSK 信号, 干扰信号表示为

$$u_j(t) = \begin{cases} U_j \cos \omega_{j1}(t), & \text{发 1} \\ U_j \cos \omega_{j0}(t), & \text{发 0} \end{cases}
 \tag{6-33}$$

忽略干扰与信号的频差, 即  $\omega_{j1} = \omega_{s1}$ ,  $\omega_{j0} = \omega_{s0}$ 。由于信号与干扰分别都有两种状态, 则分析干扰时要考虑 4 种情况。

首先看一下信号与干扰都为传号的情况。如图 6-19 所示, 此时干扰与信号的合成波经滤波器 1 输出到判决器, 而滤波 2 无信号输出。由于合成波的幅度是干扰与信号矢量和的模, 所以无论是干扰大于信号, 还是信号大于干扰, 只要干扰与信号的幅度差较大, 滤波器 1 就有较大的信号输出, 接收机将正确接收。而当干扰与信号幅度基本相同时, 一般情况下滤波器 1 有合成信号输出。但当干扰与信号反相时, 滤波器 1 的输出就非常小以至为零。此时两路滤波器的输出都为零, 择大判决时就有可能产生错误, 输出判为 1 或 0 的概率各为一半。所以干扰与信号都为传号时的错误接收概率  $P(0/1)_1$  为

$$P(0/1)_1 = \begin{cases} 0 & |U_j - U_s| \gg 0 \\ 0.5P_a & |U_j - U_s| \approx 0 \end{cases}
 \tag{6-34}$$

其中， $P_a$  为干扰可能与信号相位相反的出现概率，实际情况下  $P_a \approx 0$ ，可以忽略。

同理可得干扰与信号都为空号时的错误接收概率  $P(1/0)_0$ ，有  $P(1/0)_0 = P(0/1)_1$ ，由此可见，当干扰与信号同为传号或同为空号时，采用非相干方法进行解调时，2FSK 几乎不发生错误接收。

当信号为传号，干扰为空号时，图 6-19 中的两路滤波器同时都有输出，择大判决的结果将会是：当  $U_j < U_s$  时，滤波器 1 的输出幅度大于滤波器 2 的输出幅度，判为传号，为正确接收，即  $P(0/1)_0 = 0$ ；当  $U_j > U_s$  时，滤波器 1 的输出幅度小于滤波器 2 的输出幅度，将判为空号，为错误接收，即  $P(0/1)_0 = 1$ ；当  $U_j = U_s$  时，两路滤波器的输出幅度不相上下，判决的结果将是随机的，正确或错误的概率各为一半，即  $P(0/1)_0 = 0.5$ ，所以，此时的错误接收概率为

$$P(0/1)_0 = \begin{cases} 0, & U_j < U_s \\ 0.5, & U_j \approx U_s \\ 1, & U_j > U_s \end{cases} \quad (6-35)$$

同理，可得信号为空号、干扰为传号时的错误接收概率  $P(1/0)_1$  满足  $P(1/0)_1 = P(0/1)_0$ 。假设这四种情况发生的概率相等，则总的错误接收概率可以表示为

$$\begin{aligned} P_e &= \frac{1}{4} [P(0/1)_1 + P(1/0)_0 + P(0/1)_0 + P(1/0)_1] \\ &= \begin{cases} 0, & U_j < U_s \\ 1/4, & U_j \approx U_s \\ 1/2, & U_j > U_s \end{cases} \end{aligned} \quad (6-36)$$

对于二进制数字信号而言，当其误码率达到 1/2 时，收到的信号已无任何信息而言了，所以采用 2FSK 干扰 2FSK 信号时，当输入干信比达到 1 以上，则 2FSK 通信就被完全压制。

### 6.3.4 对扩谱抗干扰信号的干扰

适用于其他信号的干扰方程同样适用于对扩谱信号的干扰，但是扩谱信号的协作接收机能够通过压缩频谱而提供一定的“处理增益”，从而在一定程度上降低了干扰的效果。另外，扩谱信号几乎总是以数字的形式来携带信息，因此，广义上讲，扩谱信号属于数字信号的范畴，因此上文中对数字信号的干扰原则同样适用于扩谱信号。本节针对扩谱通信中最为常见的两种信号形式（DSSS 和 FHSS）进行干扰分析，其余扩谱信号的干扰分析可以参考有关资料。

#### 1. 对 DSSS 信号的干扰

接收机中的直接序列扩谱解调器的功能与发射机中的扩谱调制器功能非常相似。它对与发射机发来的伪随机序列扩谱信号进行相关运算。因此，如果施加到直接序列扩谱解调器的是窄带干扰信号，那么该信号的频谱将以与发射机中有效信号相同的扩展方式被扩展。这就使干扰信号对检测器的影响降低一个系数（该系数等于扩谱增益）。

为了便于讨论对 DSSS 通信的干扰，首先看一下干扰在 DS 接收机中的形成过程。若不考虑噪声的影响，有干扰时 DS 接收机收到的将是信号与干扰的叠加。此时，收到的信号  $r(t)$  为

$$\begin{aligned} r(t) &= s(t) + j(t) \\ &= A_c m(t) p(t) \cos \omega_c t + J(t) \cos [\omega_j t + \varphi_j(t)] \end{aligned} \quad (6-37)$$

经过解扩后输出为

$$\begin{aligned} r_o(t) &= r(t) p(t) \\ &= A_c m(t) \cos \omega_c t + J(t) p(t) \cos [\omega_j t + \varphi_j(t)] \end{aligned} \quad (6-38)$$

其中第一项为信号项，第二项为干扰项。该信号经过带宽为  $B_m$  的窄带滤波器输出后，信号项得以保留，而干扰项则被本地 PN 码  $p(t)$  扩展到比  $B_s$  宽的带宽上，使得大部分干扰不能与信号频谱重叠，而被窄带滤波器滤波。

对 DS 通信干扰进行严格的数学分析十分复杂，为简化分析，仿照 DS 系统对噪声的处理增益，引入 DS 系统对干扰的处理增益来进行对直扩信号干扰的定性分析。

首先来看一下窄带干扰的情况。由于 DS 信号带宽远大于窄带干扰的带宽，所以一般情况下，窄带干扰的全部能量都可进入宽带系统。解扩器将窄带干扰的频谱扩展到至少为  $B_s$  的宽度，使得大部分的干扰被窄带滤波器滤除掉，而只有小部分通过窄带滤波器的干扰分量才可对系统形成干扰。若窄带滤波器的带宽为  $B_m$ ，则 DS 系统对窄带干扰的处理增益为

$$G_p \geq 10 \lg \frac{B_s}{B_m} = 10 \lg N \quad (6-39)$$

单频干扰是窄带干扰的特例，单频正弦干扰进入相关解扩器，本地参考码  $p(t)$  将对它进行相关处理，使其输出为一个伪码调相波，它的功率谱密度为  $S_a(x)$  函数的形式。当正弦干扰频率与 DS 信号的伪噪声频谱中心频率相同或相近时，带内可按均匀谱处理，求出直扩系统对单频正弦干扰的处理增益为  $10 \lg N$ 。

当正弦干扰频率偏离信号伪噪声频谱中心频率较大时，特有更多的干扰成分被窄带滤波器滤除，直扩系统的处理增益会更大。可见在 DS 信号伪噪声频谱中心频率处的单频干扰，是对 DS 系统进行人为干扰的有效形式之一。

通常不相关的宽带干扰不会比窄带干扰的效果好，因为接收机前端滤波系统会对其有所抑制。即便是宽带干扰进入了接收机前端滤波系统，由于解扩器的相关处理，其带宽将会被进一步展宽，能够通过解扩后的窄带滤波器的干扰分量将很少。

DS 系统的抗干扰能力，取决于 PN 码的完善程度，如果知道了 DS 系统的 PN 码，就可采用与 DS 信号具有完全相同扩频码型的干扰信号去干扰 DS 系统，对于这种干扰信号，相关解扩后干扰与信号一样不被衰减，DS 系统对其无处理增益而言，无疑它是一种最佳的干扰样式。然而，采用此种干扰样式不仅需要掌握 DS 系统所使用的扩频码型，而且还需要干扰与信号两者精确同步，这显然是十分困难的。对于这种采用与 DS 信号相同 PN 码且能保证精确同步的干扰，我们称其为相关伪噪声扩频码干扰。它是对 DS 系统的最佳干扰样式。

当不能完全掌握被干扰目标伪码序列，但能够近似确定被干扰 DS 系统的载频及扩频码速率时，采用与目标信号具有相同载频和相同码速率、且与目标信号具有一定相关性的伪码干扰，也可获得较好的干扰效果。由于其与信号具有一定的相关性，在同步较好时，干扰与信号之间会产生较大的互相关系数，当干扰功率大到使接收机中信号对干扰的互相关峰接近或超过信号的自相关峰时，信号将被干扰有效压制。因此，采用与 DS 系统具有同样载频和码速率的 PN 码干扰也是一种对 DS 系统较有效的干扰方式。

一种实用的干扰方式是采用转发式干扰来逼近相关伪噪声扩频码干扰。假定转发式干扰

机和 DS 系统的发射机、接收机的几何位置如图 6-20 所示, 设  $r_{tr}$  为通信距离,  $r_{jr}$  和  $r_{ij}$  为干扰机到通信接收机和发射机的距离,  $t_A$  为干扰机处理时间,  $T_p$  为码元宽度,  $c$  为光速, 则有

$$\frac{(r_{ij} + r_{jr}) - r_{tr}}{c} + t_A \leq \eta T_p \tag{6-40}$$

式中,  $\eta$ 为小于 1 的数, 表示转发延时的量度。

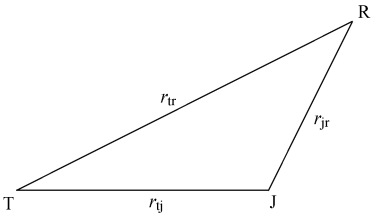


图 6-20 转发式干扰机和 DSSS 发射机、接收机的位置配置

由 PN 码的自相关特性可知,  $\eta$  越小, 转发式干扰与信号的相干函数越接近于自相关函数。在  $\eta=0$  时, 转发式干扰就是相关伪噪声扩频干扰, DS 系统对其没有处理增益; 当  $0.5 \geq \eta$  时, 处理增益损失超过一半, 因此转发式干扰是一种有效的干扰样式。

式 (6-40) 表明, 转发式干扰机必须位于以收、发信机为焦点的一个椭圆内。由于实际战术应用中最小  $r_{tr} > 5 \text{ km}$ , 而  $T_p$  一般不超过  $1 \mu\text{s}$ , 故这是一个非常狭窄的椭圆。也就是说, 只有当转发式干扰机位于 DS 系统收、发信机连线上或者附近很小的范围内时, 转发干扰才可能有效。

另外, 连续波干扰机非常简单, 相对于复杂的接收机, 它能产生的有效辐射功率更大。该技术只需将干扰功率增大到足以克服接收机处理增益的程度就能够实现有效干扰。如果信号带宽被扩展 1 000 倍, 则需要 30 dB 的干信比就可以在扩谱解调器后产生 0 dB 左右的输出干信比, 从而有效干扰扩谱信号。如果干扰机远离接收机, 这一点就难以达到, 因为我们知道, 信号强度是随着收发信机的距离而呈平方倍关系衰减。举个例子来说, 如果干扰机距离目标接收机的距离缩短为原来距离的 1/31.6, 那么其达到目标接收机的干扰信号功率将会变为原信号功率的 1 000 倍 (增加 30 dB)。

因此, 对待 DSSS 信号, 近距离干扰是非常有效的。基于这一点, 人们发明了投掷式干扰机、无人机干扰载荷和火炮投掷的干扰机等, 其目的就只有一个, 就是尽量接近目标接收机, 获得有效的强干扰信号功率。

### 2. 对 FHSS 信号的干扰

我们知道窄带干扰能够针对某个频率 (信道) 集中较大的功率实施有效的干扰, 但由于 FH 是一种躲避式的抗人为干扰的通信体制, 所以定频瞄准式窄带干扰很难对其有效干扰。

对 FH 系统采用频率跟踪式干扰, 可充分发挥窄带干扰的优势。它要求干扰能够跟踪 FH 信号频率的变化, 以保证大部分干扰功率能够进入目标信道。通常需使用压缩接收机、声光接收机、信道化接收机或多路超外差接收机等来截获所有可能的目标跳频信号, 利用高速信号处理器来识别目标信号。为了对 FH 信号实施有效的跟踪干扰, 一般要求在一个载频上的干扰时间不小于信号驻留时间的一半, 这就对干扰系统的反应速度提出了很高的要求。这种干扰对付慢跳频通信系统比较有效, 但对快速跳频通信系统干扰的效果不理想, 因为干扰系



统的反应时间很难达到对快速跳频通信干扰的要求,实施跟踪干扰需要的反应时间主要包括:信号截获时间,信号分选识别时间,干扰引导时间,以及在作用距离较远时还要考虑电波传播时延时间。信号截获时间,决定于截获 FH 信号的接收机,目前即使采用截获速度较高的压缩接收机或数字接收机,其截获时间最少也在 100 μs 左右。若要在上百个信号中分选出某一跳频信号,分选识别的处理时间也不会低于 100 μs。干扰引导时间主要取决于频率合成器的转换时间和信号建立时间,以目前频率合成器最快的转换速度计算,干扰引导所需的时间约 10 μs。电波传播的时延时间决定于干扰机到跳频收、发信机距离之和减去收、发信机之间距离之差。假设该距离差为 5 km,则电波传播引起的时延约 16 μs。将以上四部分加起来,总的时间为 226 μs。按有效干扰时间不小于信号驻留时间的一半计算,根据以上数据估算,跟踪式干扰至多可以有效干扰跳速不超过 2 000 跳/秒的跳频通信系统。实际上,目前远未达到这样的跟踪干扰速度。对宽跳频范围、大频率集、高速跳频通信系统,跟踪式干扰很难对其构成威胁。此时可采用快速频率预测跟踪瞄准式干扰来减小干扰系统的反应时间,以增加有效干扰时间。由于实现正确快速频率预测难度很大,对信号处理水平提出了更高的要求。

从 FH 通信系统的特性上看,采用全频段拦阻式干扰是对 FH 通信有效的干扰方式。这时由于跳频系统所有可能的工作频道都被干扰,系统的正常通信无法进行。但采用全频段拦阻式干扰随之而来的是干扰功率的剧增,可能会使干扰功率大到难以接受的程度。除了采用各种降低干扰功率的措施外,采用部分频段拦阻式干扰是一种行之有效的干扰方式。

当 FH 系统的全部频道数中被干扰的频道数目大到一定程度时,FH 系统的通信就不能正常工作了。即实际中一般不需要对某一 FH 系统进行全频段拦阻式干扰,一般认为,只需对其全部频道数中的大部分频道进行有效的干扰,就可获得良好的干扰效果。

下面具体分析一下对 FSK/FH 系统干扰同时击中多个信道时,FSK/FH 系统的误码率。FH 系统被干扰的情况可分为干扰同时击中预定频道及其互补频道、干扰击中预定频道而未击中互补频道和干扰未击中预定频道而击中互补频道 3 种情形。

设 FH 系统所具有的频道数为  $N$ ,干扰同时击中的频道数为  $J$ ,并设击中的干扰功率与信号功率相同。则干扰同时击中预定频道及其互补频道的概率为  $(J/N)^2$ 。此时检测判决是根据收到在预定频道及其互补频道信号功率的大小确定信号,所以只要信号与干扰迭加后的幅度小于 1,那么位于互补频道的干扰就会大于预定频道中的信号与干扰叠加而使系统产生误码。由此产生的误码率与对 2FSK 干扰双频同时击中时的情形相同,由对常规通信中 FSK 干扰的分析有

$$P_1 = \left( \frac{1}{\pi} \arccos \left( \frac{U_s}{2U_j} \right) \right) \cdot \left( \frac{J}{N} \right)^2 = \frac{1}{3} \left( \frac{J}{N} \right)^2 \quad (6-41)$$

式中,  $U_j$ 、 $U_s$  分别为干扰和信号的幅度,且假设  $U_j = U_s$ 。

干扰击中预定信道而未击中互补频道的概率为  $\frac{J}{N} \left( 1 - \frac{J}{N} \right)$ ,此时只有信号和干扰叠加后的幅度小于判决门限时才能产生误码,其概率为

$$P_2 = \frac{1}{2} P_a \cdot \frac{J}{N} \left( 1 - \frac{J}{N} \right) \quad (6-42)$$

其中,  $P_a$  为干扰可能与信号反相的概率,和式 (6-34) 中的一致,通常很小。

同理,干扰未击中预定信道而击中互补频道的概率也为  $\frac{J}{N} \left( 1 - \frac{J}{N} \right)$ ,此时系统一定产生

误码，其误码率为

$$P_3 = \frac{J}{N} \left( 1 - \frac{J}{N} \right) \quad (6-43)$$

所以，在  $J$  个干扰同时击中 FH 系统时，FH 系统的误码率为上述三式之和，即

$$\begin{aligned} P_e &= P_1 + P_2 + P_3 \\ &= \frac{1}{3} \left( \frac{J}{N} \right)^2 + \frac{1}{2} P_a \cdot \frac{J}{N} \left( 1 - \frac{J}{N} \right) + \frac{J}{N} \left( 1 - \frac{J}{N} \right) \\ &\approx \frac{1}{3} \left( \frac{J}{N} \right)^2 + P_a \cdot \frac{J}{2N} + \frac{J}{N} \quad \left( \text{当 } \frac{J}{N} \ll 1 \text{ 时} \right) \end{aligned} \quad (6-44)$$

假设 FH 系统中有十分之一的频道被干扰击中，设  $P_a = 0.01$ ，则由式 (6-44) 可得 FH 系统的误码率 0.0942，这个误码率是很高的，系统的正常工作将受到很大影响。

对 FH 系统采用转发式干扰的情形与对直扩信号干扰时相同。为了使干扰奏效，干扰机必须设置在以 FH 通信发送、接收点为焦点的一个椭圆之内，由于路径差的影响，这个椭圆是较狭窄的。一般对慢跳频通信转发干扰还是比较有效的，但对快速跳频转发干扰很难获得好的效果。采用转发干扰时除了要考虑时延影响外，通常还需要进行信号的分选，以去除掉定频信号及非目标信号。

## 思考与练习

- 6-1 试画说明通信对抗原理及通信对抗的基本内容。
- 6-2 根据香农公式说明带宽与功率互换原理。
- 6-3 请给出扩展频谱通信系统处理增益和干扰容限的计算公式，说明其意义。若  $G_p = 33 \text{ dB}$ ， $L_s = 3 \text{ dB}$ ， $(S/N)_0 = 10 \text{ dB}$ ，则输入端的干扰不超过多少时，系统仍能正常工作？
- 6-4 直接序列扩展频谱系统，又称“平均”系统或伪噪声系统，它是目前应用较广泛的一种扩展频谱系统，试画图说明其系统组成框图及工作波形示意图。
- 6-5 试画说明跳频通信系统原理方框图，快跳系统与慢跳系统有何区别？
- 6-6 跳变时间扩谱与调频扩谱有何区别？
- 6-7 试画说明 DS/Chirp 混合扩谱系统原理方框图及工作波形示意图。
- 6-8 何谓最佳干扰？若蓝方干扰机的反射功率为  $100 \text{ W}$  ( $50 \text{ dBm}$ )，其天线增益为  $10 \text{ dB}$ ，距红方接收机的距离为  $30 \text{ km}$ ；红方发射机与其接收机的距离  $10 \text{ km}$ ，发射功率为  $1 \text{ W}$  ( $30 \text{ dBm}$ )，发射天线增益为  $3 \text{ dB}$ ，接收机天线对通信信号和干扰信号的增益均为  $3 \text{ dB}$ 。试计算干信比 ( $J/S$ )。

# 第 7 章 无线信道接入方法和多址技术

无线通信系统是以信道来区分对象的，一个信道只能容纳一个用户进行通话，许多同时通话的用户互相以信道来区分，这就是多址。无线通信是一个多信道同时工作的系统，具有广播和大面积覆盖的特点。在电波覆盖区内，如何建立用户之间的无线信道的连接是多址接入方式的问题，解决多址接入问题的方法即为多址技术。

多址接入技术在无线通信中占有重要的地位，它关系到系统的构成、容量、频谱和信道利用率以及系统复杂性。

## 7.1 无线多址理论基础

通信资源（Communication Resource, CR）是指一个给定系统进行信号处理所能使用的时间和带宽，它可以用二维坐标来表示，其中横坐标表示时间、纵坐标表示频率。为了获得高效的通信系统，必须合理地有效地规划系统用户之间的资源分配，最大限度地利用时间/频率块，使系统用户能够以有效的方式共享资源，同时还要尽量不降低系统性能。

无线通信多址技术要更有效地分配通信资源，使通信系统的多个用户可以同时共享有限的无线频谱，即同时进行通信。

从无线移动通信网的构成可以知道，大部分移动通信系统都有一个或几个基站和若干个移动台。基站要和许多移动台同时通信，因此基站通常是多路的，有多个信道，而每个移动台只供一个用户使用，是单路的。许多用户同时通信时，以不同的信道分隔，防止相互干扰。各用户信号通过在射频频道上的复用建立各自的信道，以实现双边通信的连接。因此，多址接入或多址连接是多用户无线通信网中按用户地址进行连接的问题，即解决众多用户如何高效共享给定通信资源的问题。

### 7.1.1 无线双工通信方式

在无线通信系统中，为了使通信的双方可以同时发送和接收信号，必须采用双工通信方式，即每一个用户信道是由两个信道所组成，一个信道用来发送信号，另一个信道用来接收信号。这可以通过频域技术或时域技术来实现。

#### 1. 频分双工

两个信道通过频率来区分，这种技术称为频分双工（Frequency Division Duplex, FDD）。FDD 技术为每个用户提供了两个确定的频段：前向频段和反向频段，如图 7-1 所示。在无线移动通信中，前向频段（也称为前向信道）提供从基站到移动用户的信号传输信道（下行信道），反向频段（也称为反向信道）提供从移动用户到基站的信号传输信道（上行信道）。

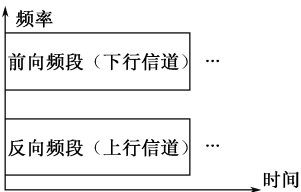


图 7-1 频分双工

在 FDD 中，任何双工信道实际上都是由两个单工信道所组成，利用在用户和基站里的称为双工器的设备，允许同时在双工信道上进行无线发射和接收。前向信道和反向信道的频率分隔在整个系统中是固定的，为了尽量减少每个用户信道上前向波段与反向波段之间的相互干扰，应在通信系统的频谱范围内，使频率分隔尽可能大一些。

FDD 适用于为每个用户提供单个无线频率信道的无线通信系统。由于每个用户通信时，需要同时地发送和接收相差大于 100 dB 的无线信号，所以必须谨慎地分配用于前向信道和反向信道的频率，使其与占用这两个波段之间频谱的其他用户保持协调。此外，频率的分配和选择必须考虑到降低射频（RF）设备成本。

2. 时分双工

两个信道通过时隙来区分，这种技术称为时分双工（Time Division Duplex, TDD）。TDD 技术是用时间而不是用频率来提供前向信道和反向信道，如图 7-2 所示。如果前向时隙和反向时隙之间的时间间隔很小，那么对于信号的发送和接收，在用户看起来就是同时的。由于 TDD 是在同一个频率信道上通过时间分隔来提供前向信道和反向信道，所以不需要双工器，从而简化了用户设备。

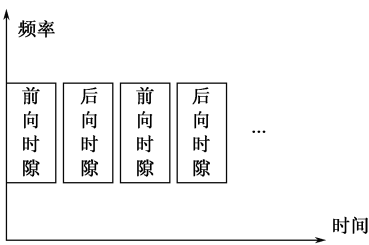


图 7-2 时分双工

TDD 可以使每个用户在同一频段上的不同时刻内发送和接收信号，发送信号的时隙与接收信号的时隙之间应保留一段非常小的时间间隔作为保护间隔，以防止发送信号与接收信号之间的相互冲突。此外，通信的双方还要严格地保持时间上的同步，以保证能够正确地接收信号。

7.1.2 无线信号分割原理

在不考虑天线极化和空间问题，无线多址技术与信道多路复用技术一样，它们的理论基础都是信号分割原理。不同的是多路复用技术的目的在于区分多路信号，通常在中频或基带实现，而多址接入技术的目的在于区分多个无线用户，必须在射频实现。

无线电射频信号可以表达为频率、时间和码型的函数，可写成

$$s(c, f, t) = c(t) \cdot s(f, t) \tag{7-1}$$

式中， $c(t)$ 是码型函数； $s(f, t)$ 是时间和频率的函数。

信号分割的依据在于信号之间的差别，这种差别可以是频率参量上的，可以是时间参量上的，也可以是码型结构上的。

当仅以传输信号的载波频率的不同来区分信道建立多址连接时，称为频分多址（Frequency Division Multiple Access, FDMA）；当仅以传输信号存在时间的不同来区分信道建立多址连接时，称为时分多址（Time Division Multiple Access, TDMA）；当仅以传输信号码型的不同来区分信道建立多址连接时，称为码分多址（Code Division Multiple Access, CDMA）。这是 3 种最基本的通信多址方式，下面分别予以介绍。

# 7.2 频分多址技术

## 7.2.1 FDMA 基本原理

在整个通信领域，不论无线通信还是有线通信，FDMA 是最经典的多址技术，在通信电缆、卫星通路和各种无线通信网中，很多使用频分多址。

频分多址用频率作为用户信号的分割参量，将给定的频谱资源划分为若干互不交叠的频段（或称频道、信道），这些信道按照一定的规则分配给系统用户，一般是分配给每个用户一个

唯一的频段（信道）。在该用户通信的整个过程中，其他用户不能共享这一频段。

频分多址是以频率来区分信道的，多个频道在频率轴上严格分开，但在时间和空间上是重叠的，此时，“信道”一词的含义即为“频道”。

在频分双工频分多址工作方式（FDMA/FDD）中，每个用户分配两个频道，其中频率较高的频道用做前向信道，频率较低的频道用做反向信道，即接收频道和发送频道不同。

如果用频率、时间和代码作为三维空间的 3 个坐标，则 FDMA/FDD 数据通信系统在这个坐标系中的位置如图 7-3 所示，它表示系统的每个用户由不同的频道所区分，但可以在同一时间、用同一代码进行通信。

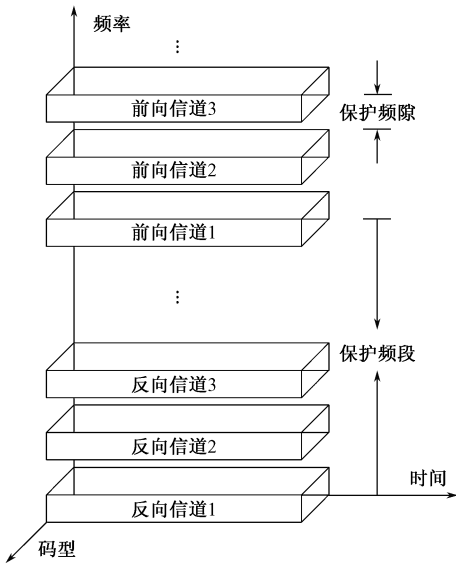


图 7-3 频分双工频分多址工作方式（FDMA/FDD）

为了防止各用户收、发信号相互干扰，各用户收、发频道之间通常都要留有一段间隔频段，称为保护频段，此间隔必须大于一定的数值，例如，在 800 MHz 和 900 MHz 频段，收、发频率间隔通常为 45 MHz。此外，为避免系统频率漂移造成正向频道或反向频道之间的重叠，在各频道之间设有保护频隙。

## 7.2.2 FDMA 系统中的干扰问题

FDMA/FDD 系统是按照频率段来划分信道的。每个用户在一对信道（前向信道和反向信道）中通信。如果有其他信号的成分落入系统用户的频率段内时，将造成对用户信号的干扰。在蜂窝小区内的基站与移动台系统中，主要干扰有互调干扰和邻道干扰。在频率集重复使用的蜂窝系统中，还存在着同频道干扰。

### 1. 互调干扰

互调干扰是指系统内某个用户的发射机（或接收机）由于非线性器件产生的各种组合频率成分落入本身的前向信道（或反向信道）频率段内而造成的对其自身有用信号的干扰。当

干扰的强度（功率）足够大时，将会影响用户信号的正常发送和接收。克服互调干扰的主要办法：一是尽可能提高系统的线性度，减少发射机互调和接收机互调，消除产生互调干扰的条件；二是选用无互调的频率集，这涉及到 FDMA 蜂窝系统的频率规划问题。

## 2. 邻道干扰

邻道干扰是指系统内相邻用户的发射机（或接收机）由于非线性器件产生的各种组合频率成分落入本用户信道的频率段内而造成的对本用户有用信号的干扰。当邻道干扰功率足够大时，将对有用信号造成损害。克服邻道干扰的主要方法：一是加大频道间的隔离度，即增大保护频段，这也涉及到 FDMA 系统的频率规划问题；二是严格规定收发信机的技术指标，即限制发射机的寄生辐射和规定接收机的中频选择性。

## 3. 同频道干扰

同频道干扰是指相同频率信号之间的相互干扰。对系统中的某一用户而言，是指同一蜂窝系统的邻近小区中的同频道信号造成的干扰。它与蜂窝结构和频率规划密切相关。为了减少同频道干扰，需要合理地选择蜂窝结构和科学地进行频率规划，具体为在系统设计中同频道干扰因子  $Q$  的选择：

$$Q = \frac{D}{R} \tag{7-2}$$

式中， $D$  为同信道小区的距离（即同频小区的距离）； $R$  为小区半径。

### 7.2.3 FDMA 系统的特点

模拟信号和数字信号都可采用频分多址方式传输，但在模拟无线通信系统中，采用 FDMA 方式是唯一的选择。该方式有如下特点：

（1）单路单载频。每个频道只传送一路信息，每个频道带宽并应满足所传输信号的带宽要求。为了在有限的频谱中增加信道数量，希望载频带宽越窄越好。FDMA 信道的相对带宽较窄（25 kHz 或 30 kHz），每个信道的每一载波仅支持一个电路连接。

（2）连续传输。系统分配给移动台和基站一对 FDMA 信道，它们利用此频道通信直到结束。因此在 FDMA 系统中，分配好用户信道后，移动台越区切换时，必须瞬时中断传输数十至数百毫秒，以便把通信从某一频率切换到另一频率上去。对于语音，瞬时中断问题不大，对于数据传输则将带来数据的丢失。

（3）是频道受限和干扰受限的系统。FDMA 系统中载波带宽与单个信道一一对应，在接收设备中的窄带滤波器可使指定的信道信号通过，并滤除其他频率的信号，从而可拟制邻道干扰、互调干扰和同频干扰。

（4）在 FDMA 方式中，每个信道只传送一路数据信号，信号速率低，码元宽度与平均延迟相比较是很大的，这就意味着由码间干扰引起的误码极小，因此在窄带 FDMA 系统中无须进行均衡。

（5）需要周密的频率计划，频率分配工作复杂。

（6）频率利用率低，系统容量小。

FDMA 系统可以同时支持的信道（用户）数可用下列公式计算：

$$N = \frac{B_s - 2B_p}{B_c} \quad (7-3)$$

式 (7-3) 中：  $B_s$  为系统带宽，  $B_p$  为在分配频谱时的保护带宽，  $B_c$  为信道带宽。

## 7.3 时分多址技术

### 7.3.1 TDMA 基本原理

时分多址系统以时间作为信号分割的参量来区分信道的，它把时间划分为叫做时隙的时间小段， $N$  个时隙组成一帧，无论是时隙与时隙之间还是帧与帧之间在时间轴上必须互不重叠。然后根据一定的时隙分配原则，例在每一帧中固定位置周期性重复出现的一系列离散的时隙组成一个信道，系统的每一个用户占用一个这样信道。

时分多址方式中，时间轴按时隙上严格分割，时隙间无保护时间，并在频率轴上是重叠的，此时，“信道”一词的含义即为“时隙”。

在 TDMA 中，各个用户在每帧中只能在规定的时隙内向基站发射信号（突发信号），在满足定时和同步的条件下，基站可以在相应的时隙中接收到相应用户的信号而互不干扰。同时，基站发向各个用户的信号都按顺序安排在规定的时间隙中传输，各个用户只要在规定的时间隙内接收，就能从时分多路复用（TDM）的信号中接收到发给它的信号。

如果用频率、时间和代码作为三维空间的三个坐标，则 TDMA 数据通信系统在这个坐标系中的位置如图 7-4 所示，它表示系统的每个用户由不同的时隙所区分，但可以在同一频段、用同一代码进行通信。

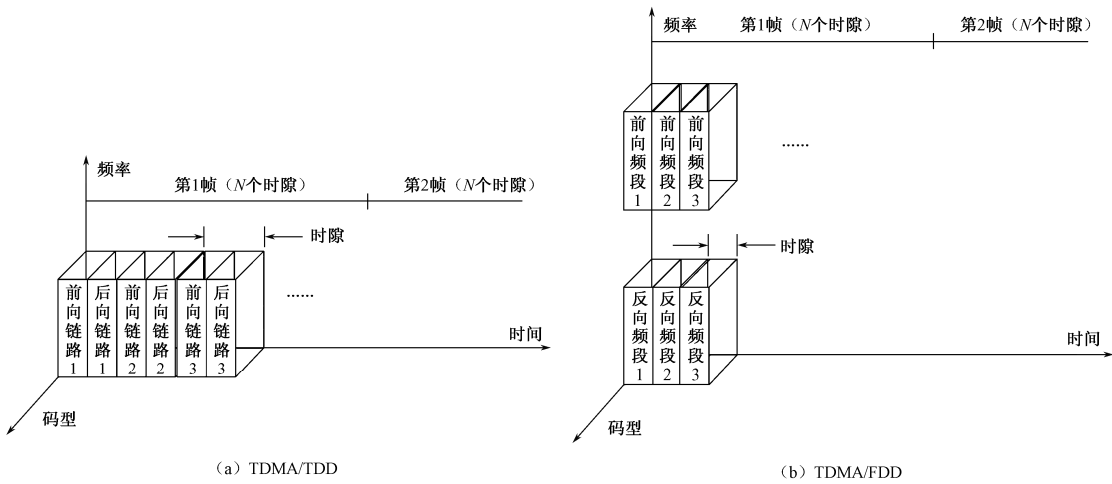


图 7-4 时分双工与频分双工的时分多址工作方式

在时分双工/时分多址工作方式（TDMA/TDD）中，帧结构中的每个时隙的一半用于前向链路，而另一半用于反向链路。

在频分双工/时分多址工作方式（TDMA/FDD）中，其前向频段和反向频段有一个完全相同或相似的帧结构，前者用于前向传送，后者用于反向传送。

### 7.3.2 TDMA 的帧结构

TDMA 帧是 TDMA 系统接收、处理和传输信息的基本单元，它是由若干个时隙所组成，不同的用户周期性地占有一系列特定的时隙来传送自己的信号，但在每一个 TDMA 帧中一个用户只占有一个特定的时隙。不同通信系统的帧长度和帧结构是不一样的。典型的帧长在几毫秒到几十毫秒之间，例如，GSM 系统的帧长为 4.6 ms（每帧由 8 个时隙组成）。TDMA/TDD 帧结构示意图如图 7-5 所示。

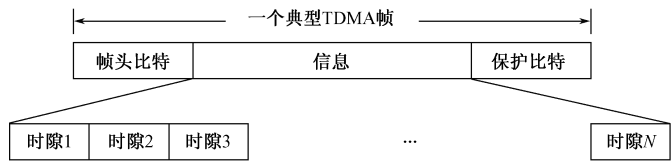


图 7-5 TDMA/TDD 帧结构

从图 7-5 中可以看出，一个 TDMA 帧由头比特、信息段和保护比特所构成，其中头比特包含了基站和用户用来确认彼此的地址信息和同步信息，信息段包含了各用户时隙，而保护比特则用来保证帧与帧之间的同步。

由于 TDMA 系统是以时间分割来区分不同的信道、通信双方只允许在规定的时隙中发送信号和接收信号，所以在时间上同步是 TDMA 通信系统正常工作的前提条件。因而 TDMA 通信系统必须在严格的位同步、帧同步和系统同步的条件下进行工作。如果通信系统的接收设备采用相干解调，则发射机与接收机之间还必须满足载波同步。在进行 TDMA 帧结构设计时，必须满足位同步、帧同步和系统同步的要求。

### 7.3.3 TDMA 系统的特点

TDMA 系统有以下特点：

(1) TDMA 系统通过分配给每个用户一个互不重叠的时隙，使  $N$  个用户可以共享同一个载波频道，所以它的频带利用率高，系统容量大。

(2)  $N$  个时分信道共用同一个载波频道，占据相同带宽，只需一部收发信机，所以互调干扰小，基站设备的复杂性减小。

(3) 由于 TDMA 系统是在不同的时隙内来发射和接收信号的，因此不需要双工器。就是对于使用了 FDD 双工技术的 TDMA 系统，也可以通过合理地设计 TDMA/FDD 系统中前向频段和反向频段的帧结构来避免使用双工器，从而简化了系统的设备。

(4) 越区切换简单。由于在 TDMA 系统中移动台是不连续的突发式传输，所以越区切换处理对一个用户单元来说是很简单的，可以安排在无信息传输时（对某个特定用户而言是空闲时隙）进行，因而没有必要中断信息的传输，即使正在传输数据也不会因越区切换而发生数据丢失的现象。

(5) TDMA 系统的各个用户的数据发射不是连续的，而是分组发送的，这样当用户发射机（在大多数时间里）不用时可以关机，从而大大地降低了电池消耗。

(6) TDMA 系统可以实现按需动态分配时隙，即在一帧中分配给不同的用户不同数目的



时隙，通过基于优先权重重新分配时隙的方式，按照不同用户的不同要求来提供带宽。

(7) 由于 TDMA 发射被时隙化了，因此要求接收机与每一个数据分组必须保持同步，使 TDMA 系统需要较高的同步开销。此外，为了将不同的用户分开，还需要设立一定宽度的保护时隙，这就又占用了一部分时间资源，所以 TDMA 系统相对于 FDMA 系统有更大的系统开销。

(8) 发射信号速率一般比 FDMA 信道的发射速率高得多，并且随着时分信道数目  $N$  的增加而提高，如果达到 100 kb/s 以上，码间串扰就会变大，必须采用自适应均衡，用以补偿传输失真。

(9) 在 TDMA 系统中，为了充分利用时间资源，应把保护时间压缩到最小。但是为了缩短保护时间而把时隙边缘的发射信号加以明显抑制，将使发射信号的频谱扩展并会导致对相邻信道的干扰。

### 7.3.4 TDMA 系统的帧效率与系统容量

#### 1. 帧效率

TDMA 系统的帧效率是指一帧中用户数据比特数占总比特数的百分比，此处的用户数据可以包括原始数据和信道编码，因此一个系统的原始终端用户的效率通常小于  $\eta_f$ 。帧效率  $\eta_f$  的计算公式如下：

$$\eta_f = \frac{n_u}{n} \times 100 \% = \left( 1 - \frac{n_s}{n} \right) \times 100 \% \tag{7-4}$$

式中， $n_s$  为每一帧中系统开销的比特数， $n_u$  为用户数据的比特数， $n$  为每一帧的总比特数（等于一帧中用户数据的比特数与系统开销的比特数之和）。

#### 2. 系统容量

在实际应用中，通常是把 TDMA 方式与 FDMA 方式结合起来使用，这样可以大大地提高系统的容量。此时系统容量（用户数）的计算公式如下：

$$N = \frac{m(B_s - 2B_p)}{B_c} \tag{7-5}$$

式中， $B_s$  为系统带宽， $B_p$  为在分配频谱时的保护带宽， $B_c$  为信道带宽， $m$  为每一个频率信道所含的时隙数。

该系统容量（用户数）的计算公式也可以用来计算纯 TDMA 系统容量，此时的频率信道数为 1， $N = m$ 。

## 7.4 码分多址技术

### 7.4.1 CDMA 基本原理

#### 1. 正交结构函数（序列）概念

因正交结构函数（序列）概念是码分多址技术的基础，所以我们首先予以介绍。

m 序列、Gold 序列等伪噪声随机序列（Pseudo Noise, PN），以及沃尔什（Walsh）函数

等为常用的正交结构函数。以图 7-6 所示典型沃尔什函数的波形为例，说明其特性。

沃尔什码序列的波形是定义在半开区间 $[0, 1)$ 的矩形波族，每个矩形波有一个编号  $n(n=0, 1, 2, 3, \cdots)$ ，矩形波幅度的取值为+1 或-1，规定起始时矩形波的取值为+1，然后在+1 与-1 之间变化，变化的次数（+1 变-1 与-1 变+1 的次数之和） $m=n$ ，在+1 或-1 上持续的时间可以相等，也可以不相等（不相等时较长的持续时间为较短的持续时间的两倍），并将编号为  $n$  的沃尔什函数表示为  $Wal(n, t)$ 。

一般信息 0 可用沃尔什函数的正向波形表示（调制），信息 1 可用沃尔什函数的反向波形表示（调制），数字信息比特通过如此调制后再进行传输。

沃尔什函数基本性质中最重要的性质是正交性，即

$$\int_0^1 Wal(i, t) \cdot Wal(j, t) dt = \begin{cases} 1, & i = j \\ 0, & i \neq j \end{cases} \quad (i, j = 0, 1, 2, \cdots) \quad (7-6)$$

也就是说，沃尔什函数在同步时是完全正交的，只有用户的地址码波形与接收到的信号波形码结构完全相同时，才能从中取出有用信息，否则不可以。

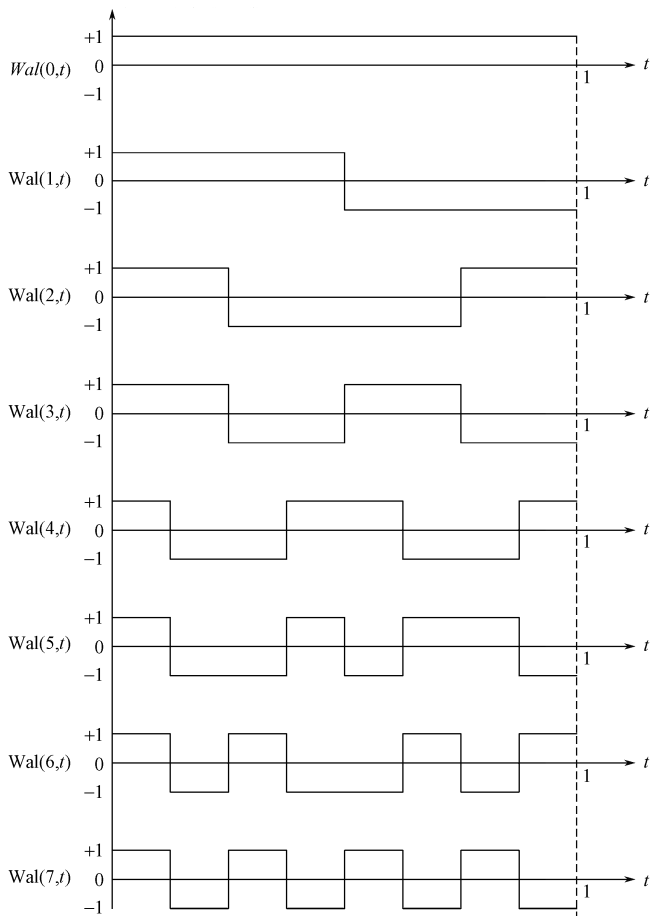


图 7-6 典型沃尔什函数波形举例

2. 码分多址原理

码分多址系统以码型结构作为信号分割的参量，它为每个用户分配了各自特定的地址码。系统的各用户互不相关的、相互正交的地址码调制其所发送的信号，在接收端利用码型的正交性，通过地址识别（相关检测）从混合信号中选出相应的信号。

在 CDMA 方式中，不同的用户传输信息所用的信号不是根据频率或时隙的不同来区分的，而是用各不相同的码型结构序列来区分的。如果从频域或时域来观察，多个 CDMA 信号是互相重叠的。接收机用相关器可以在多个 CDMA 信号中选出其中使用预定码型的信号，而其他使用不同码型的信号不能被解调。它们的存在类似于在信道中引入了噪声和干扰（称为多址干扰），各码型之间的互相关性越小，多址干扰就越小。

CDMA 系统无论传送何种信息的信道，都是靠采用不同的码型来区分的，所以，此时“信道”一词的含义为“码型”或“码道”。

如图 7-7 所示，在频分双工/码分多址工作方式（CDMA/FDD）中，每一个用户分配有一个地址码，这些码型信号相互正交（即码型互不重叠），并在同一载波上传输，其用户的前向信道和反向信道采用频率划分实现双工通信。

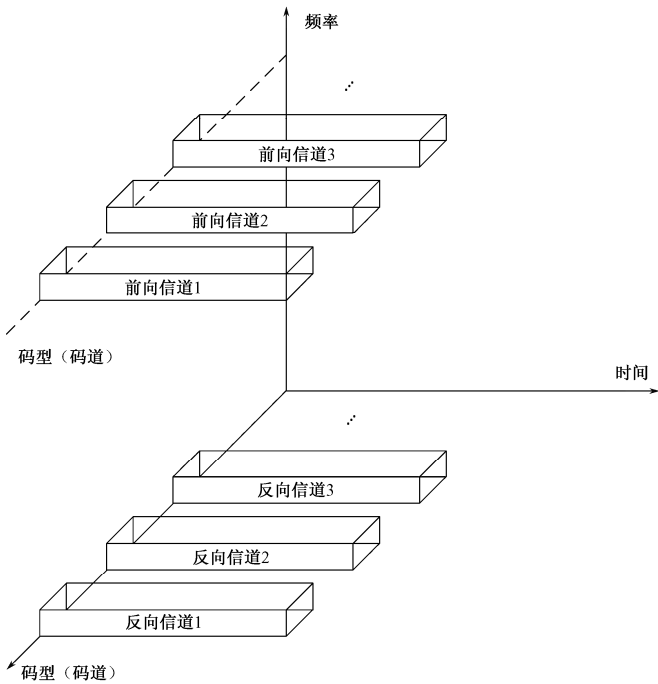


图 7-7 频分双工/码分多址工作方式（CDMA/FDD）

在码分多址工作方式，为了实现双工通信，我们可以采用频分双工技术，也可以采用时分双工技术。

7.4.2 CDMA 系统的特点

CDMA 系统有以下特点：

- （1）频率共享，CDMA 系统可以实现多用户在同一时间内使用同一频率进行各自的通信

而不会相互干扰。

(2) 通信容量大, 由于对一个 CDMA 系统用户而言, 其他用户信号相当于噪声, 这样增加 CDMA 系统中的用户数目会线性增加噪声背景, 使系统的性能下降, 但不会中断通信, 所以 CDMA 系统具有软容量特性, 对用户数目没有绝对限制。这就是说 CDMA 是干扰限制性系统, 干扰的增加会降低系统的容量, 而干扰的减少会提高系统的容量, 因此可以利用一些抗干扰技术来提高系统容量。通过分析计算, 在相同条件下, CDMA 的系统容量大约是 TDMA 系统的 5 倍, 是 FDMA 系统的 12 倍。

(3) 抗多径衰落, 我们知道, 当频谱带宽比信道的相关带宽大时, 那么固有的频率分集将具有减少小尺度衰落的作用。由于 CDMA 是扩频系统, 其信号被扩展在一个较宽的频谱上, 从而可以减小多径衰落。

(4) 更好的接收效果, 在 CDMA 系统中, 信道数据速率很高。因此码片时长很短, 通常比信道的时延扩展小得多, 因为 PN 序列有低的自相关性, 所以超过一个码片时延的多径将被认为是噪声, 受到接收机的自然抑制; 另一方面, 如果采用分集接收最大合并比技术, 可获得最佳的抗多径衰落效果, 提高接收的可靠性。

(5) 平滑的软切换, CDMA 移动通信系统中所有小区使用相同的频率, 所以它可以用宏空间分集中来进行软切换, 使越区切换得以平滑地完成。当移动台处于小区边缘时, 同时有两个或两个以上的基站向该移动台发送相同的信号, 移动台的分集接收机能同时接收合并这些信号, 此时处于宏分集状态。当某一基站的信号强于当前基站信号且稳定后, 移动台会自动切换到该基站的控制上去, 这种切换可以在通信的过程中平滑完成, 称为软切换。软切换由移动交换中心 (MSC) 来执行, 它可以同时监视来自两个以上基站的特定用户信号, 选择任意时刻信号最好的一个, 而不用切换频率。

(6) 低信号功率谱密度, 在 CDMA 移动通信系统中, 信号功率被扩展到比自身频带宽度宽百倍以上频带范围内, 因而其功率谱密度大大降低。由此可得到两方面的好处, 一是具有较强的抗窄带干扰能力; 二是对窄带系统的干扰很小, 有可能与其他系统共用频段, 使有限的频谱资源得到更充分的使用。

### 7.4.3 CDMA 系统的两个问题

虽然 CDMA 系统具有较多的优越性, 但也存在着两个重要的问题, 一个是自干扰问题, 另一个是“远一近”效应问题。

#### 1. 自干扰问题

由于 CDMA 系统中不同的用户采用的扩频序列 (地址码) 不是完全正交的, 在同步状态下, 各用户序列的互相关系数虽然不为零但比较小, 在非同步状态下, 各用户序列的互相关系数不但不为零, 有时还比较大。这一点与 FDMA 和 TDMA 是不同的, FDMA 具有合理的保护频隙, TDMA 具有合理的保护时隙, 接收信号近似保持正交; 而 CDMA 对这种正交性是不能保证的。这种扩频码集的非零互相关系数引起的各用户之间的相互干扰被称为多址干扰 (Multiple Access Interference, MAI), 在异步传输信道以及多径传播环境中多址干扰将更为严重。由于这种干扰是系统本身产生的, 所以称为自干扰。解决自干扰问题的根本办法是找到在同步状态下和非同步状态下序列的互相关系数均为零的数字序列。

## 2. 远近效应问题

如果 CDMA 移动通信系统中不同的用户都以相同的功率发射信号的话,那么离基站近的用户接收功率就会高于离基站远的用户的接收功率。这样在不同位置的用户,其信号在基站的接收状况将会不同。即使各用户到基站的距离相等,各用户信道上的不同衰落也会使到达基站的信号各不相同。如果期望用户与基站的距离比干扰用户与基站的距离远得多,那么干扰用户的信号在基站的接收功率就会比期望用户信号的接收功率大得多(最大可以相差 80 dB)。

在同步 CDMA 系统中,接收功率的不同不会产生不良影响,因为不同用户信号之间是严格正交的;在非同步 CDMA 系统中,接收功率的不同有可能产生严重的影响,因为此时不同用户的非同步扩频波形不再是严格正交的,从而对弱信号有着明显的抑制作用,会使弱信号的接收性能很差甚至无法通信。这种现象被称为“远近效应”。

为了解决远近效应问题,在大多数 CDMA 实际系统中使用功率控制。蜂窝系统中由基站来提供功率控制,以保证在基站覆盖区内的每一个用户给基站提供相同功率的信号。这就解决了由于一个邻近用户的信号过强而覆盖了远处用户信号的问题。基站的功率控制是通过快速抽样每一个移动终端的无线信号强度指示(Radio Signal Strength Indication, RSSI)来实现的。尽管在每一个小区内使用功率控制,但小区外的移动终端还会产生不在接收基站控制内的干扰。

## 7.5 扩频及混合多址技术

目前扩频及混合多址技术主要有 3 种:直接序列扩频多址(Direct Sequence Spread Spectrum Multiple Access, DS-SSMA)、跳频多址(Frequency Hopping Multiple Access, FHMA)和混合扩频多址(Hybrid Spread Spectrum Multiple Access, HSSMA)。

直接序列扩频多址也叫做码分多址(CDMA),是最基本的多址技术之一,根据扩频后的频谱(或扩频序列码的长短)分为窄带 CDMA 和宽带 CDMA,本章前面已有所讨论,这里不再介绍,下面仅介绍 FHMA 和 HSSMA。

### 7.5.1 跳频多址

单纯跳频多址(FHMA)系统与频分多址系统一样,仍仅以频率作为用户信号的分割参量,但不同的是在跳频多址系统中,分配给每个用户的频带(或载波)不是固定不变,而是在系统信道带宽范围内按伪随机 PN 代码方式不断变化,也就是说,系统所有用户占用的频带(或载波)是在不断地按伪随机方式重新分配。图 7-8 说明了跳频多址系统的工作方式。

假设跳频系统的有效信道带宽被划分为  $n$  个频带(或载波),则最多可以有  $n$  个用户占用这  $n$  个频带(或载波),但各个用户在各时间段(时隙)所占用的频带(或载波)是不同的,所以,用户数据将在不同的载波上发射,并且任一个发射组的瞬时带宽都比整个信道扩展带宽小得多。

在 FHMA 接收机中,用当地产生的 PN 代码来使接收机的瞬时载波与发送机同步,以便正确地接收跳频信号。

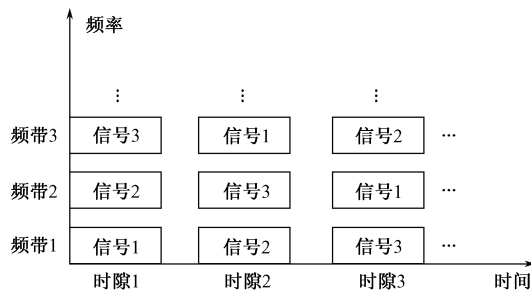


图 7-8 跳频多址系统频谱的再分配

如果 FHMA 系统中各个用户的载波变化速率(跳频速率)大于系统传输信号的码元速率,那么该系统就被称为快跳频系统;如果载波变化速率小于或等于码元速率,那么该系统就被称为慢跳频系统。

一个快跳频系统可以被认为是使用频率分集的 FDMA 系统。FHMA 系统经常使用能量效率高的恒包络调制,并用廉价的接收机来提供 FHMA 的非相干检测。

跳频系统具有良好的保密性能,尤其是可供选择频带(或载波)的数量比较多时,因为想窃取信息的接收机并不知道频带(或载波)是怎样随机改变的,所以不能很快地调谐到它希望监听的动态频带(或载波)上并与之保持同步。

跳频系统中偶尔会在某些频段出现深度衰落,可以用纠错编码和交织技术来保证跳频信号不受衰落的影响。纠错编码和交织技术也可用来防止碰撞(两个或多个用户同时在同一频段上发射称为碰撞)的影响。

## 7.5.2 混合扩频多址

在频率、时间或码型的信号参量中,用两个以上信号参量来分割用户的技术称为混合多址技术,常用的有频分/码分、时分/码分、跳频/码分、时分/跳频等混合多址技术。由于这些混合多址方式中均使用扩频技术,所以常称为混合扩频多址(HSSMA)技术。这些技术各具优点,下面就简单讨论其原理。

### 1. 频分/码分混合多址技术

频分/码分多址(FDMA/CDMA)又称频分 CDMA(FCDMA),是频分多址技术与码分多址技术相结合而形成的一种混合扩频技术。

首先将系统的有效宽带频谱划分为(频分)若干个子频谱,这些子频谱不是直接分配给各用户,而是分配给各窄带 CDMA 系统。也就是说以码分多址为基础,占用有效宽带频谱中的一个子频谱,形成一个窄带 CDMA 系统,一个一个频谱分开的窄带 CDMA 系统,构成了频分/码分混合多址方式。图 7-9 说明了各用户采用频分双工工作时,用户的区分采用频分/码分混合多址方式。

频分 CDMA 系统有以下优点:一是整个系统的有效带宽可以不连续;二是可以根据不同用户的不同要求分配其在不同的子频谱上;三是整个系统的容量就是所有窄带 CDMA 系统的容量之和。

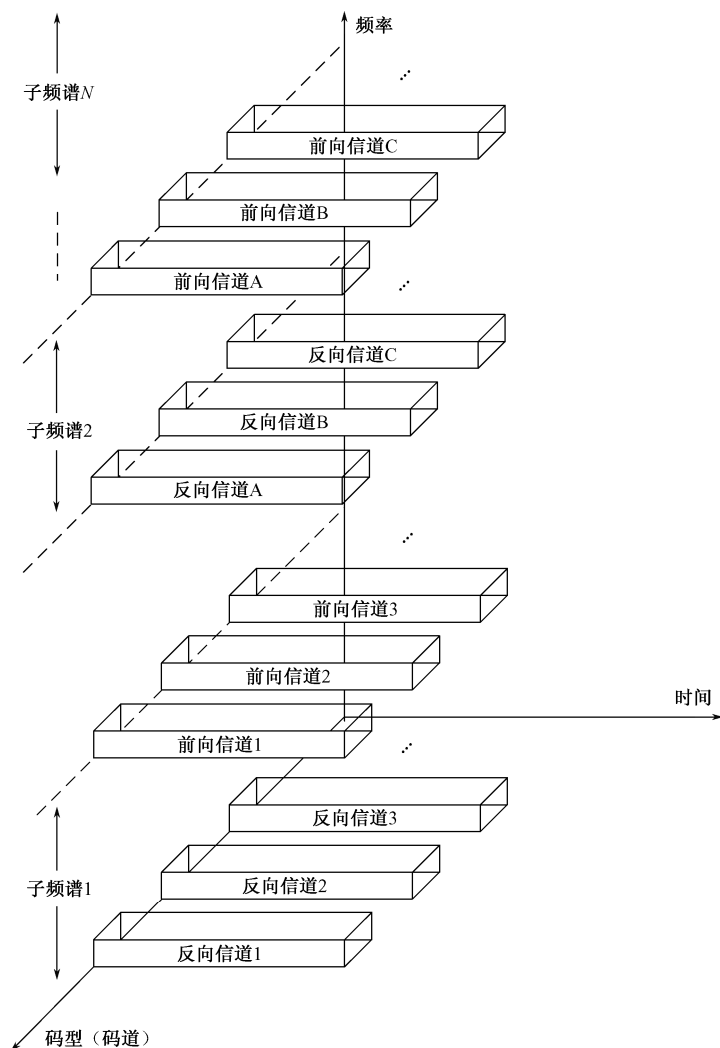


图 7-9 频分双工时的频分/码分混合多址方式示意图

## 2. 时分/码分混合多址技术

时分/码分多址 (TDMA/CDMA) 又称时分 CDMA (TCDMA)，是时分多址技术与码分多址技术相结合而形成的一种混合扩频技术，如图 7-10 所示。

在 TCDMA 系统中，一个码分地址（扩频代码）不是直接分配给一个特定用户的，而是分配给一个特定的小区用户群，在小区内，给每个用户再分配一个特定的时隙（时分）。因此在任意时刻，每一小区只有一个 CDMA 用户在发射和接收信号。当发生从一个小区到另一个小区的切换时，该用户的扩频代码就变成新小区的扩频代码。

时分 CDMA 系统的优点是：由于在一个小区内的任一时刻只有一个用户在发射和接收信号，避免了远近效应。

## 3. 跳频/码分混合多址技术

跳频/码分多址 (FHMA/CDMA) 又称跳频 CDMA (FHCDMA)，是跳频多址技术与码

分多址技术相结合而形成的一种混合扩频技术。它以码分多址为基础，占用有效宽带频谱中的一个子频谱，形成一个窄带 CDMA 系统，但是这个窄带 CDMA 系统所占用的子频谱不是固定的，而是以伪随机方式在有效宽带频谱范围内跳变（跳频）。

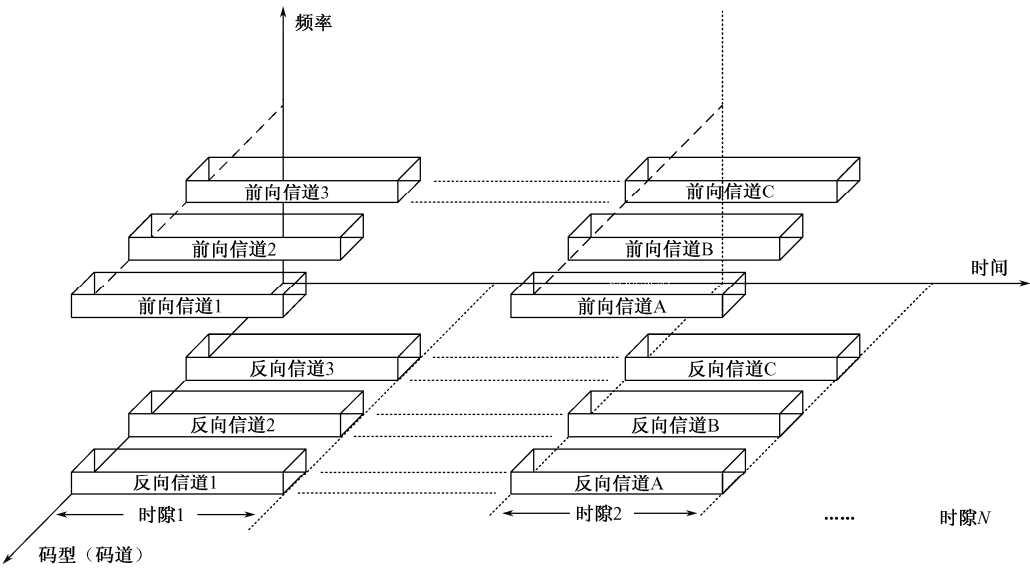


图 7-10 频分双工时的时分/码分混合多址方式示意图

跳频 CDMA 系统的优点是避免了远近效应，然而这种混合系统不适用于软切换处理，因为很难使跳频 CDMA 基站接收机和多路跳频信号同步。

跳频 CDMA 与频分 CDMA 的区别是：FHCDMA 中有多个窄带 CDMA 系统，它们各自占用 FHCDMA 系统有效宽带频谱范围内若干个子频谱中的一个，但不是固定不变而是随时间变化，即随时间不断地以伪随机方式重新分配子频谱；FCDMA 也有多个窄带 CDMA 系统，但是每个窄带 CDMA 系统固定地占用 FCDMA 系统有效宽带频谱范围内的一个子频谱，不随时间变化。

4. 时分/跳频混合多址技术

时分/跳频多址（TDMA/FHMA）又称时分 FHMA（TFHMA），是时分多址技术与跳频多址技术相结合而形成的一种混合扩频技术。它以跳频多址为基础，但每个跳频地址不是分配给一个用户而是分配给若干个用户，然后再把系统占用某一频段的时间划分为若干个时隙分配给这若干个用户（时分），从而构成时分 FHMA 系统。

时分/跳频多址系统有以下优点：一是虽然每个用户也是周期性地占用一个时隙（TDMA 帧），但该用户在本时隙采用一个发射频率，在下一个时隙就会跳到一个新的发射频率（跳频），因此避免了在一个特定信道上的严重多径衰落或碰撞事件；二是如果采取措施使两个互相干扰的基站发射机在不同频率和不同时间发射，可以避免邻近小区的同信道干扰；三是采用这种混合扩频技术可以成倍增加系统容量。

GSM 标准已经采用此项技术。在 GSM 标准中，预先定义了跳频序列，并且允许用户在指定小区的特定频率上跳频。



## 7.6 正交频分复用多址接入

正交频分复用（OFDM）本身是一种调制技术，但它可以很容易地与多种多址接入技术相结合，为多个用户同时提供接入服务。常用的多址接入方式有时分多址（TDMA）、频分多址（FDMA）和码分多址（CDMA），OFDM 都可以与它们结合，分别构成 OFDM-TDMA，OFDM-FDMA 和 OFDM-CDMA，下面就来介绍一下这 3 种技术。

### 7.6.1 OFDM-TDMA

在正交频分复用时分多址（OFDM-TDMA）系统中，信息的传送是在时域上按帧来进行的，每个时间帧包含多个时隙，每个时隙的宽度等于 1 个 OFDM 符号的时间长度，有信息要传送的用户根据各自的需求可以占用 1 个或多个 OFDM 符号。每个用户在信息传送期间，可占用所有的系统带宽，即该用户的信息可以在 OFDM 的所有子载波上进行分配。

### 7.6.2 OFDM-CDMA

OFDM 与 CDMA 扩频技术相结合的方法可分为两类：频域扩频和时域扩频。频域扩频通常称为 Multicarrier-CDMA（简称 MC-CDMA）。时域扩频有两种不同的构成方法：分别称为 Multicarrier DS-CDMA（简称 MC-DS-CDMA）和 Multitone CDMA（简称 MT-CDMA），下面介绍一下三种不同的 OFDM-CDMA。

#### 1. MC-CDMA

MC-CDMA 是最早提出的 OFDM 与 CDMA 相结合的方案。在此方案中，每个信息符号先经过与扩频序列各位相乘，相乘后对应于不同码片的信号分别调制到不同的子载波上，若扩频序列长度为  $L$ ，信息符号则分别调制到  $L$  个子载波上，调制方式可采用二进制相移键控（BPSK）。如假定 OFDM 系统共有  $L$  个子载波，则 CDMA 系统的扩频增益等于  $L$ 。

在直接序列扩频（DS-CDMA）系统中，信息在许多时间码片上用同一载波频率进行发送，而在 MC-CDMA 系统中，信息是在许多载波频率码片上同时进行发送的。

可见，DS-CDMA 与 MC-CDMA 系统之间有“时间/频率”的对应关系：MC-CDMA 把信息同时调制在不同载波频率分量上（频率码片），接收时对频率码片进行分集接收；DS-CDMA 把信息同时调制在不同的时隙（时间码片）上，但是使用同一载波频率，接收时对时间码片进行分集接收。

#### 2. MC-DS-CDMA

在 MC-DS-CDMA 方案中，输入信息比特首先进行串/并变换，被分配到并行支路上，然后，各支路上的信息符号分别用长度为  $L$  的扩频码进行直接序列扩频，扩频后的信号再分别用各自的载波进行 BPSK 调制，调制后的信号进行求和后发送。

#### 3. MT-CDMA

在 MT-CDMA 方案中，输入的信息符号首先经过串/并变换，调制到不同的载波上，以

形成 OFDM 信号，OFDM 的符号周期为  $T_s$ 。然后再经过长度为  $L$  的扩频码扩频，扩频后每个子载波的带宽扩展为  $L/T_s$ ，而相邻子载波的间隔仍然保持以前的  $1/T_s$ 。MT-CDMA 一般采用较长的扩频序列，比 DS-CDMA 能容纳更多用户。

### 7.6.3 OFDM-FDMA

OFDM-FDMA 在许多文献中又被称为 OFDMA。这种多址接入方案与传统的频分复用 (FDMA) 很类似，它通过为每个用户提供部分可用子载波的方法来实现多用户接入。与传统 FDMA 的不同之处在于，OFDMA 方法不需要在各个用户频率之间采用保护频段去区分不同的用户。

OFDMA 接入方案的优势之一是可以很容易地引入跳频技术，即在每个时隙中，可以根据跳频图样来选择每个用户所使用的子载波频率。每个用户使用不同的跳频图样进行跳频，可以把 OFDM 系统变化成为跳频 CDMA 系统，从而可以利用跳频的优点为 OFDM 系统带来好处。

与直扩 CDMA 或者 MC-CDMA 相比，跳频 OFDMA 的最大好处在于可以为小区内的多个用户设计正交跳频图样，从而可以较容易地消除小区内的干扰。

## 7.7 空间与极化多址方式

在无线多址技术中，除根据信号的频率、时间或码型的参量中来分割用户的技术外，还可以根据信号发射时极化方式和天线辐射波束的空间特征区分用户，并称为极化区分多址和空分多址方式，下面就简单讨论这些多址原理。

### 7.7.1 空分多址 (SDMA)

空分多址 (Space Division Multiple Access, SDMA) 也称为多波束频带再利用 (Multiple Beam Frequency Reuse)，它是以空间特征作为用户信号的分割参量，目前利用最多也是最明显的空间特征就是用户的位置，利用电磁波传播的特点可以使不同地域的用户在同一时间使用同一频率实现互不干扰的通信。

具体地讲，可以利用定向天线或窄波束天线使电磁波按一定方向辐射并局限在窄波束范围内，不同波束范围可以使用相同的频率；也可以通过控制发射功率使电磁波作用在有限距离内，在电磁波作用范围以外的地域仍可使用相同的频率，从而实现以空间区分不同的用户。图 7-11 为定向窄波束辐射时的空分多址方式示意图。

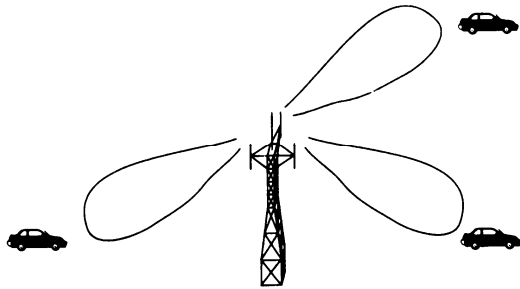


图 7-11 空分多址方式的工作示意图

空分多址是一种比较早期的多址方式，在频率资源管理上早已使用。蜂窝移动通信就是由于充分运用了这种多址方式，才能用有限的频谱构成大容量的通信系统，不过在蜂窝移动通信中把这种技术称为频率再用技术。

此外，一个完善的自适应式天线系统，可以实现在小区域内不受其他用户干扰的唯一信道，即提供最理想的 SDMA。同时，也可以在小区域内搜索用户的多个多径分量，并且以最理想的方式组合它们，来收集从每个用户发来的所有有效信号能量，从而有效地克服多径干扰和同信道干扰。尽管上述理想情况是不可实现的，因为它需要无限多个阵元。但采用适当数目的阵元，还是可以获得较大的系统增益。

卫星通信中采用窄波束天线实现空分多址，也提高了频谱的利用率。但由于波束的分辨率是非常有限的，即使卫星天线采用了阵列处理技术后波束的分辨率有较大的提高，也还是不能满足实际应用的要求，所以空分多址通常与其他多址方式综合运用。

激光束的方向性非常好、散射非常小，一束激光从地球传播到月球（地球到月球的距离为 38.44 万公里），所覆盖面积的直径只有几公里至几十公里，如果从距地面几百公里的卫星上发射到地面，所覆盖面积的直径只有几米至几十米，其分辨率非常高。随着空间激光通信的深入研究，必将为空分多址方式的应用开辟更加广阔的前景。

近年来，人们发现空间特征不仅仅是位置，在技术飞速发展的今天，一些过去认为无法使用的空间特征现在正逐步被人们利用，形成以智能天线为基础的新一代空分多址方式。

7.7.2 极化区分多址（PDMA）

极化区分多址（Polarization Division Multiple Access, PDMA）也称为双极化频带再利用（Dual Polarization Frequency Reuse）。

天线向周围空间辐射电磁波，电磁波由电场  $E$  和磁场  $H$  构成，人们规定电场  $E$  的方向就是电磁波的极化方向，如图 7-12 所示，其中电场矢量方向与地面平行时称为水平极化，电场矢量方向与地面垂直时称为垂直极化。

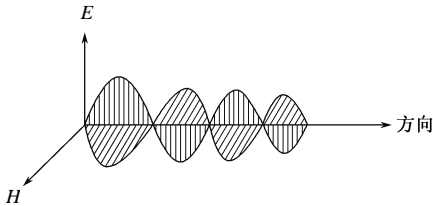


图 7-12 电磁波传播示意图

极化区分多址系统的理论依据是：水平方向极化天线只能接收到水平极化电磁波，不能接收垂直极化电磁波；垂直方向极化天线只能接收到垂直极化电磁波，不能接收水平极化电磁波。因水平和垂直波正交，当用户发射机采用水平或垂直极化天线发射电磁波时，与之相同极化的接收机天线，才能较好地接收该种极化用户的电磁波，与之不相同极化的接收机天线，不能接收该电磁波，即根据用户的不同极化方式进行区分，实现多址通信。

例如在卫星通信中，地球上的同一地区在同一频段上电磁波采用水平与垂直极化，可以有两路信号接入卫星，从而使频带利用率翻了一番。图 7-13 说明了卫星通信极化区分多址系统的工作方式。

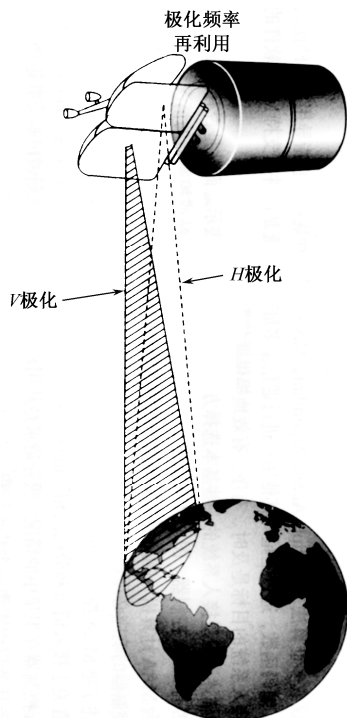


图 7-13 卫星通信极化区分多址方式的工作示意图

## 思考与练习

- 7-1 试说明无线双工通信方式与无线通信多址技术的区别。
- 7-2 频分多址是以信号的何种参量区分信道？并画图说明频分多址技术有什么特点。
- 7-3 画图说明时分双工时与频分双工时的时分多址工作方式的区别。
- 7-4 画图说明频分双工码分多址工作方式，并说明特点。
- 7-5 如图 7-14 所示，说明图中所采用的是何种工作方式，有何优缺点。
- 7-6 什么是跳频/码分混合多址工作方式，画图说明。
- 7-7 在无线多址技术中，除根据信号的频率、时间或码型等参量来分割用户的技术外，还可以根据哪些特征区分用户？

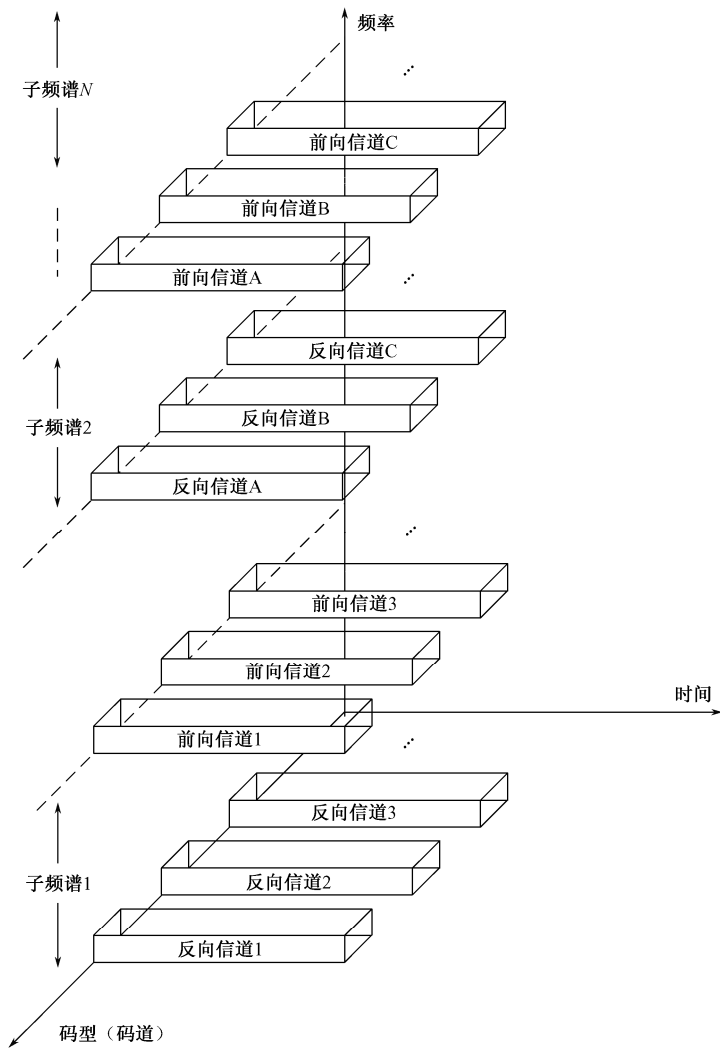


图 7-14

# 第 8 章 超宽带 (UWB) 通信技术

超宽带 (Ultra-Wide Band, UWB) 技术是利用超宽频带的电波进行高速无线通信的技术。从时域上讲,超宽带系统有别于传统的通信系统。一般的通信系统是通过发送射频载波进行信号调制,而 UWB 是利用起、落点的时域脉冲(几十纳秒)直接实现调制。超宽带的传输把调制信息过程放在一个非常宽的频带上进行,而且以这一过程中所持续的时间来决定带宽所占据的频率范围。

## 8.1 超宽带基本概念

### 8.1.1 超宽带定义及规范

无线超宽带通信技术是目前无线通信领域较先进的技术之一。这项技术曾被美国军方秘密用于二战,当时这项技术称为冲激无线电 (Impulse Radio),主要是利用脉冲通信减少干扰和阻塞,增强通信的准确性、可靠性和隐蔽性,随后关于这项技术的研究工作也主要局限于军方,直到 20 世纪 90 年代,“超宽带”这一术语由美国国防部首先提出,并应用于超宽带通信、超宽带导航、超宽带雷达、超宽带微波炸弹等。目前,它逐步转入民用阶段,并在无线电、音、视频和数据传输及家用设备领域内得到迅速发展。

随着电子技术的飞速发展,人们对脉冲无线电技术的认识也更加清楚,2002 年美国联邦通信委员会 (FCC) 发布了针对超宽带的报告和规范,规定只要一个信号的相对带宽 $\Delta f/f_0$ 大于等于 25 %,或者-10 dB 绝对带宽大于 500 MHz,则这个信号就是超宽带信号。其中,相对带宽
$$\frac{\Delta f}{f_0} = \frac{f_H - f_L}{f_0} \frac{f_H - f_L}{(f_H + f_L)/2}$$
 $f_H$  为信号最高频率,  $f_L$  为信号最低频率,  $f_0$  为信号中间频率。

为了避免民用的超宽带系统对已有的无线通信系统 (GPS、蜂窝移动通信、802.11n) 产生干扰, FCC 规范根据三类用途的超宽带通信系统可能产生的干扰,对它们的频谱使用范围和功率辐射进行了严格和具体规定。

三类用途的超宽带通信系统为:

- (1) 包括探地雷达 (GPR)、穿墙雷达的成像系统、监视器以及医疗成像设备;
- (2) 车载雷达系统;
- (3) 通信和测量系统。

FCC 对不同用途的超宽带设备频谱使用范围的规定如下:

- (1) 探地雷达与墙壁成像系统: 低于 960 MHz 或 3.1~10.6 GHz;
- (2) 墙壁穿透成像系统: 低于 960 MHz 或 1.99~10.6 GHz;
- (3) 监视系统: 1.99~10.6 GHz;
- (4) 医疗系统、通信和测量系统: 3.1~10.6 GHz;
- (5) 车载雷达: 22~29 GHz, 另外, 中心频率和最高辐射电平点的频率必须大于 24.075 GHz。

FCC 对超宽带设备的功率辐射限制以 EIRP (Effective Isotropic Radiated Power) 指标给

出。所谓 EIRP，即有效全向辐射功率，是一个天线的输入功率与某个指定方向天线增益的乘积相对全向天线的值。

### 8.1.2 超宽带通信与现有通信的比较

由于超宽带无线电是采用脉冲机制的实现方式，因此它与传统的无线电的发射信号是不同的。

#### 1. 信号时域、频域波形不一样

超宽带信号波形一般为冲激脉冲信号，其频谱宽、平均功率低，现有通信信号一般多为正弦载波，其频谱窄、平均功率高。

从超宽带信号的时域上看，其信号的时域特性也与传统无线电信号的时域特性有着明显的差异，主要表现在以下两个方面：

超宽带信号在时域上是没有载波的，发射的信号是一串低占空比的窄脉冲，而传统通信体制下，发射的信号都需要载波作为信息的载体；

超宽带信号在时域上呈现非连续特性（由于低占空比的因素），而传统通信体制的发射信号一般在时域上呈现波形连续特性。

超宽带信号的频谱，无论从频谱幅度和频谱宽度上都与传统无线电信号的频谱有较大的区别。超宽带直接发射冲激脉冲串，不需要传统无线电所需的载波，它可以认为是不需要混频的，是基带信号，而从另一个方面看，由于超宽带直接发射基带的脉冲串，空间的频谱结构又与传统的无线电并无本质区别，因此可以认为超宽带信号即是基带信号，又是射频信号。而超宽带所具有的空间宏观频谱结构与其所采用的基带脉冲波形紧密联系在一起的，而其细微空间频谱结构是由采用的具体调制方式所决定的。

#### 2. 通信器材不一样

超宽带通信所有元器件都必须具有超宽带性能，现有通信元器件为常规器件，可以容易在市场上购置。

#### 3. 检测手段不同

超宽带信号一般采用频域检测，测量谱密度。现有通信信号采用时域检测，测量峰值功率。

#### 4. 通信体制可以不同

UWB 通信可以被分类为一种扩频技术，但又与使用特定载波的常规扩频技术不同，它发送的是波形不变的窄脉冲，这种脉冲持续时间非常短（一般为纳秒级），波形中有过零点。根据傅里叶变换原理，时域内信号持续时间越短，相应的频域上占据的频带就越宽，信号能量在频谱内分布的也就越广，进而实现扩频。

在 UWB 通信系统中，为实现多用户同时通信（即多址通信），采用了跳时多址（Time Hopping Multiple Access, THMA）方式，即用伪随机码改变脉冲在时间轴上出现的位置，利用不同的伪随机码来区分不同的用户，只有拥有相同伪随机码的用户才能相互通信。

## 5. 优于“蓝牙”技术

超宽带通信适应复杂环境（城市、室内），通信距离为 10 m~50 km（“蓝牙” <100 m），可用于室内、通信和大范围蜂窝组网，且传输速率比“蓝牙”高，更适应多媒体业务，抗干扰能力比“蓝牙”强。

实现超宽带通信的首要任务是产生 UWB 信号。从本质上看，UWB 是发射和接收超短电磁脉冲的技术。可使用不同的方式来产生和接收这些信号以及对传输信息进行编码，这些脉冲可以单独发射或成组发射，并可根据脉冲幅度、相位和脉冲位置对信息进行编码（调制）。

### 8.1.3 超宽带通信的特点

从上述对超宽带信号时域与频域的描述中可以看出，超宽带信号在时域、频域两个方面都与传统无线电信号有着较大差异，正因为这些显著的差异，直接导致了超宽带信号有着传统无线电信号不具备的特性或优点，主要表现在以下几个方面。

#### 1. 发射机和接收机相对简单

不需要载波，且发送和接收设备简单，是 UWB 技术的重要特点。由于 UWB 信号是一些超短时的脉冲，其频率很高，故发射器可直接用脉冲激励天线，且不需要功放与混频器；同时在接收端也不需要中频处理，因此，必然会使发射机和接收机的结构简单化。

#### 2. 功耗低

信息论中关于信息容量的香农（Shannon）公式为

$$C = B \log_2 \left( 1 + \frac{S}{N} \right) \quad (8-1)$$

式中， $C$  为信道容量（用传输速率度量）， $B$  为信号频带宽度， $S$  为信号功率， $N$  为白噪声功率。式（8-1）说明，在给定的传输速率  $C$  不变的条件下，频带宽度  $B$  和信噪比  $S/N$  是可以互换的，这可通过增加频带宽度的方法，来降低系统对信噪比的要求，即当发射信号功率很小（如 50~70 mW）时，也可以保证在适当传输距离上实现正常通信。

正因如此，UWB 技术的平均发射功率很低，在短距离应用中，UWB 发射机的发射功率通常可以低于 1 mW，从而大大延长了电源的供电时间。FCC 规定，UWB 的发射功率谱密度必须低于美国发射噪音规定值 -41.3 dBm/MHz，因此，从理论上来说，相对于其他通信系统，UWB 信号所产生的干扰仅相当于一宽带白噪声。民用 UWB 设备的功率一般是传统移动电话所需功率的 1/100 左右，是蓝牙设备所需功率的 1/20 左右，而且军用 UWB 电台的耗电也很低。因此，UWB 设备在电池寿命和电磁辐射上，相对于传统无线设备有着很大的优越性。

#### 3. 传输速率高

超宽带脉冲信号和系统的频带极宽，一般在几百 MHz 到几 GHz，即脉冲码元速率可以达到 10 Gbps，这是一般通信体制无法达到的高速率。目前，超宽带通信已经可以在很低的信噪比门限下实现大于 100 Mb/s 的可靠高速无线传输，且其进一步的目标是 500 Mb/s 和 1 Gb/s。



一个相同作用范围的超宽带通信系统,其速率可以达到无线局域网 802.11b 系统的 10 倍以上,蓝牙 (Bluetooth) 系统的 100 倍以上。

4. 隐蔽性与安全性好

隐蔽性好、安全性高是 UWB 技术的另一重要特点。由于 UWB 信号的带宽很宽,且发射功率很低,这必然使其具有低截获能力。另外,超宽带还采用了扩频技术,接收端必须在知道发射端扩频码的条件下才能解调出发送的数据信息,因而提高了安全特性。

5. 距离分辨率高

利用通信电波来回传输的时间长短,确定距离的计算公式为

$$d = \frac{\tau}{2} c \tag{8-2}$$

即传输的时间 $\tau=10^{-9} \sim 10^{-12}$ s 时,距离分辨率在厘米以内。

由于常规无线通信的射频信号大多为连续信号或其持续时间远大于多径传播时间,所以,在这些通信系统中,多径传播效应限制了通信质量和数据传输速率。而在 UWB 系统中,从时域角度看,超宽带系统采用的脉冲宽度为几纳秒的窄信号,脉冲持续时间极短,而且具有极低的占空比,脉冲重复周期远远大于脉冲宽度,且远大于多径时延,这就使得所有时延大于脉冲宽度而小于脉冲周期的多径分量都可以明确地分辨出来,因此,超宽带信号具有很强的多径分辨能力。

6. 能够克服多径干扰

由于 UWB 信号的带宽极宽,脉冲宽带一般为 ns 量级~ps 量级 ( $10^{-9} \sim 10^{-12}$ ),只要反射信号延时大于 2 ns,就能被相关接收机滤除,因而容易克服多径干扰。

另外,从频域的角度分析,所以信号在传输过程中一定会出现频率选择性衰落现象。然而,正是因为极宽的带宽,多径衰落只在某些频点出现,从整体上考虑,衰落掉的能量只是信号总能量很小的一部分。

7. 频带利用率高及信道容量大

从时域看,超宽带通信是对超窄脉冲进行调制,脉冲波形有梯形波、钟形波、锯齿波,脉冲可调参数有宽度、重复频率、上升下降时间等,即可供调制的参数比正弦型载波要多得多,在不占有现有频率资源的情况下,可带来一种全新通信方式。

8. 抗干扰能力强

在常用的直接序列扩频系统中,采用高速率的伪随机码和低速率的信息数据进行相关运算来实现扩频,扩频后的频谱一般为几十兆赫。超宽带通信系统采用的跳时扩频方式,是利用窄脉冲信号本身的频谱特性进行扩频,扩展后的频谱为几千兆赫兹,是一般扩频系统的一百倍。因此,在具有相同信息码速率的情况下,超宽带通信系统比一般扩频系统的处理增益大 20 dB 左右。另外,超宽带通信脉冲型载波可调参数多,差错率低,有不同于常规的通信体制,所以具有更强的抗干扰能力,特别适用于军事信息对抗。

## 9. 穿透能力强

超窄脉冲的频谱非常丰富，它能穿透冰层、海水、丛林、大地等，从而可以辨别隐藏在物体背后的目标，实现与深水潜艇的通信等：

$$\text{穿透深度 } \delta = \sqrt{\frac{2}{\omega \mu \sigma}} \quad (8-3)$$

全球卫星定位系统只能定出可视范围内的目标位置，而超宽带脉冲极强的穿透能力可以定出地下等重要目标，特别是军事目标，也可以对存放在货架上的贵重物品进行跟踪监控。

## 8.2 超宽带信号的基本调制方式

实现超宽带通信的首要任务是产生 UWB 信号。从本质上看，UWB 是发射和接收超短电磁脉冲的技术，可使用不同的方式来产生和接收这些信号以及对传输信息进行编码，这些脉冲可以单独发射或成组发射，并可根据脉冲幅度、相位和脉冲位置对信息进行编码（调制）。

超宽带无线通信信号的调制方式随着超宽带通信技术研究的不断深入，从早期的基于冲激无线电的调制方式，发展到现今的基于频域处理的调制方式，可谓日新月异。在关于超宽带调制众多的方式中，本节重点讨论超宽带基于传统的冲激无线电的调制方式。目前较成熟的 UWB 通信信号有两种：时间调制（Time-Modulated, TM）UWB 和直接序列相位编码（Direct Sequence Phase Coded, DSC）UWB。

TM-UWB 采用瞬时开关技术来产生窄脉冲。可以将能量扩展到很宽的频带内。窄脉冲由专用宽带天线，以每秒几十兆至几百兆赫的高速率进行发射，这些脉冲在时间上以随机或伪随机间隔进行分布，进而可在时域和频域上产生多个噪声编码信道和一个类似噪声的信号。对脉冲进行时间编码后就形成多个信道，实现多址通信。具体实现有基于跳时体制的脉冲位置调制（UWB TH-PPM）、基于跳时体制的脉冲幅度调制（UWB TH-PAM）。

DSC-UWB 采用一种以 Gb/s 量级的速率发射的高占空比脉冲的相位编码序列。脉冲序列以数百 Mb/s 的速率对数据进行编码（调制），多个编码脉冲表示一个比特，编码增益能提供抗多径干扰能力，在短距离范围内，DSC-UWB 能提供极高的数据速率。具体实现有基于脉冲的直接序列调制（DS-UWB）。

超宽带调制方式中的基于频域处理的方式，本书不打算进行介绍，当然，超宽带无线通信的调制方式还有很多，有兴趣的读者可参考相关书籍，进行更详细的了解和学习。

### 8.2.1 基于跳时体制的脉冲位置调制（UWB TH-PPM）

由于 UWB 利用非连续的脉冲串作为信息的载体，这与传统无线通信技术所利用的信息载体有着本质的区别，因此在调制方式中，基于脉冲的调制方式成为 UWB 调制方式中的首选。TH-PPM 是基于传统的冲激无线电的调制方式中提出最早和应用最广泛的一种调制方式。早在 1993 年 R. A. Scholtz 在军事通信会议上便提出了 TH-PPM 调制的最初原形，后有各国的学者和研究机构不断对其进行改进和完善，使其从最初的基于点对点的通信调制样式逐渐发展到适用于多用户，并将 UWB 系统与传统通信用户的共存问题进行了综合考虑。

## 1. 时域信号结构

一般的讲 UWB TH-PPM 调制可分为二进制 UWB TH-PPM 调制和多进制 UWB TH-PPM 正交调制两种。下面我们主要介绍二进制 UWB TH-PPM 调制方法，该调制方法的数学表达式如下：

$$S_{\text{TH-PPM}}(t) = \sum_{j=-\infty}^{j=+\infty} \sqrt{E_p} W(t - jT_f - c_j T_c - \delta d_{\lfloor j/N_S \rfloor}) \quad (8-4)$$

式中， $E_p$  为 UWB 信号所调制的单个脉冲的能量； $T_f$  为脉冲的平均重复时间； $c_j$  是跳时码中的第  $j$  个码，一般取  $0 < c_j \leq N_{\max}$ ； $T_c$  是跳时码的单位时移，也是跳时码对第  $j$  个脉冲的附加时移，跳时序列的周期为  $N_p$ ； $\delta$  为 2-PPM 的调制指数； $d_{\lfloor j/N_S \rfloor} \in (0,1)$  是待发送的二进制信息数据， $N_S$  为每个信息比特含有的脉冲数， $\lfloor \cdot \rfloor$  为取整运算； $W(t)$  代表发射的极窄脉冲波形。

脉冲的带宽一般约为该脉冲持续时间的倒数，从式 (8-4) 中可以看出，UWB 系统的每个比特的持续时间为  $T_b = N_S T_f$ ，则信息速率为  $R_b = 1/T_b$ 。通常，在进行 UWB TH-PPM 调制的设计时，取  $N_{\max} T_c + \delta < T_f$ ，这是为了保证 UWB 发射信号中每帧中的脉冲经过 TH 码调制后不会发生混叠，这有利于 UWB 系统削弱信道中多径的影响，然而若  $N_{\max} T_c + \delta$  取的太小（对减弱多径影响有利），当系统中的用户较多时，容易发生多用户共址冲突，因此不利于 UWB 的多址系统。UWB 发射的脉冲波形，可取高斯形（高斯脉冲的各阶微分波形和组合波形）和瑞利形等，由于高斯形脉冲有接近于零的直流分量，便于天线发射，因此本书采用前者，又由于 UWB 接收机所接收的脉冲波形通常是 UWB 发射机发射波形的一阶微分，所以此处取高斯形脉冲的一阶微分（主要是由于发射天线的影），其波形函数表达式为

$$W(t) = 2\sqrt{e} A_p \pi t f_c e^{-2(\pi t f_c)^2} \quad (8-5)$$

在式 (8-5) 中，幅度  $A_p$  的取值应使该脉冲具有单位能量，该脉冲的时间宽度定义为：该时长的脉冲能量应占有总能量的 99.99 %。采用一阶高斯脉冲作为 UWB 系统发射脉冲波形的 2-TH-PPM 调制波形如图 8-1 所示。

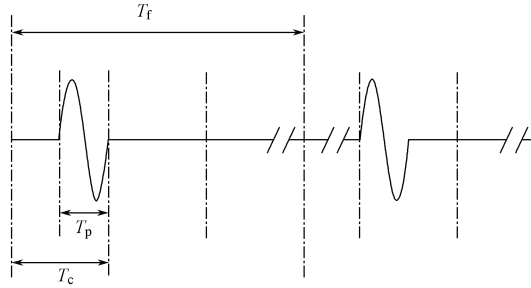


图 8-1 UWB 2-TH-PPM 系统发射脉冲波形示意图（采用一阶高斯脉冲）

从图 8-1 中可以看出，UWB TH-PPM 调制中，一个信息比特一般包含多个脉冲帧，而脉冲帧的重复频率一般被称为重复脉冲频率（Pulse Repeation Frequency, PRF），然而由于 TH 码的调制，脉冲的间隔是非固定的，因此一般认为  $1/T_f$  为平均的脉冲重复频率，该参数对 UWB 系统具有重要的意义，该参数不仅影响 UWB 系统的数据传输速率，而且直接关系到 UWB 系统对共存通信系统的干扰效果。实际应用中，一般取  $T_f$  等于整数倍  $T_c$ ，该整数是由跳时码的集合中的最大值确定的。总起来说，二进制 UWB TH-PPM 调制方式是实际应用中

使用最为广泛的 TH-PPM 调制方式。

2. 收发信机的结构原理

由以上 UWB 信号模型的特点，可以得到相应的发射机与接收机系统构成示意图，如图 8-2 所示。

发射时，时钟发生器产生一定重复周期的脉冲序列，然后根据用户要传输的数据信息和表示该用户地址的伪随机码，通过可编程延时器对系统时钟进行精确的延时，延时后的脉冲序列驱动脉冲产生电路，从而形成一定脉冲形状和规律的脉冲序列，然后耦合到超宽带天线发射出去。具有皮秒（ $10^{-12}\text{s}$ ）级分辨率的延时器由集成电路实现，它是 TM-UWB 系统实现中的关键技术部件。

在接收端，超宽带天线接收的信号将送到相关器的一个输入端，相关器的另一个输入端加入一个本地产生的、与发端同步的、经用户伪随机码调制的脉冲序列一起经过相关器中的相乘、积分和采样保持运算，然后产生一个对用户地址信息经过分离的信号，该信号中仅含用户传输信息及其他干扰。然后对该信号进行解调运算，即根据发端的调制方式对每个脉冲进行判决，恢复出所传输的信息。接收机中的基带处理包括捕获和跟踪电路，可用于准确锁定时间编码序列。

从图 8-2 中可以看出，与传统无线收、发信机的结构相比，该方案下的 UWB 收发信机结构相对比较简单。

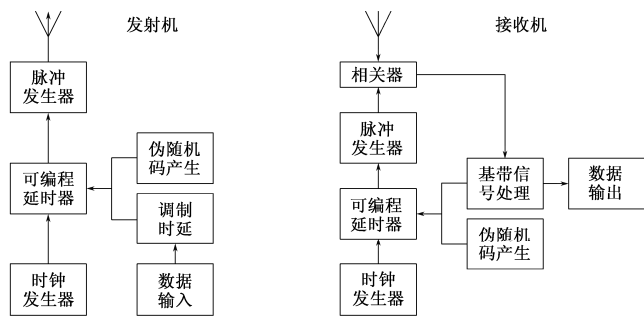


图 8-2 UWB 发射机与接收机系统构成示意图

8.2.2 基于跳时体制的脉冲幅度调制（UWB TH-PAM）

TH-PAM 调制也是超宽带无线通信中的一种典型调制方式，它是利用脉冲的幅度参量来进行信息的传输，也可分为二进制 TH-PAM 调制和多进制 TH-PAM 调制两种，这里只简要介绍二进制 TH-PAM 调制方式。

二进制 UWB TH-PAM 调制信号的一般表达式如下：

$$S_{\text{TH-PAM}}(t) = \sum_{j=-\infty}^{j=+\infty} A_{[d/N_s]} \sqrt{E_p} W(t - jT_f - c_j T_c) \tag{8-6}$$

式中， $A_{[d/N_s]} \in (-1, +1)$ ，其他参数的意义与式（8-4）中介绍的相同。采用一阶高斯脉冲作为 UWB 系统发射脉冲波形的二进制 UWB TH-PAM 波形如图 8-3 所示。

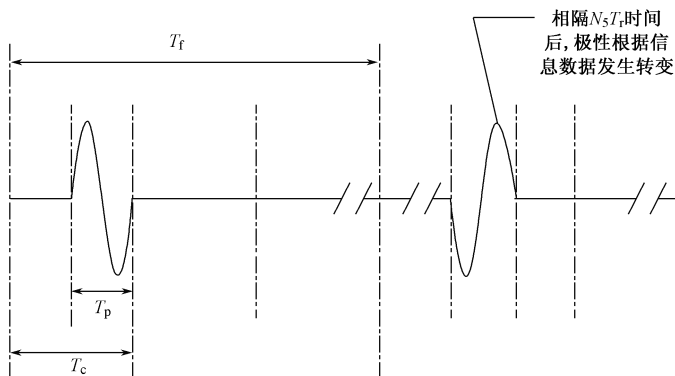


图 8-3 UWB 二进制 TH-PAM 系统发射脉冲波形示意图（采用一阶高斯脉冲）

从图 8-3 中可以看出，二进制 TH-PAM 调制方式与二进制 TH-PPM 调制方式的信号，主要区别是：二进制 TH-PPM 调制中的脉冲在一帧中的位置是根据跳时码和此时的信息数据共同决定的，而二进制 TH-PAM 调制中的脉冲在一帧中的位置仅取决于跳时码，与信息无关。另外二进制 TH-PPM 调制中脉冲的极性不会发生改变，即与信息无关；而二进制 TH-PAM 调制中脉冲的极性是根据信息数据发生改变的。

多进制 TH-PAM 调制方式中，脉冲的幅度大小和极性都可能发生改变，这主要取决于发送信息与脉冲幅度的映射关系。当该调制方式中不采用跳时码时，脉冲的位置是等间隔出现的，极性则是根据信息发生变化。

### 8.2.3 基于脉冲的直接序列调制（DS-UWB）

基于脉冲的直接序列调制是超宽带无线通信的另一种调制方式，也是美国电气和电子工程师学会（Institute of Electrical & Electronic Engineers, IEEE）关于 UWB 技术的几个竞选标准之一。它将传统扩频思想与 UWB 技术很好地融合在了一起，使各自的优势都得到了体现。经典的 DS-UWB 信号的信号波形表达式为：

$$S_{\text{DS-UWB}}(t) = \sum_{j=-\infty}^{\infty} \sum_{n=0}^{N_s-1} d_j g_n \sqrt{E_p} W(t - jT_f - nT_c) \quad (8-7)$$

式中， $d_j \in (+1, -1)$  是所调制的信息数据； $g_n \in (+1, -1)$  是用户的伪随机序列中对应 UWB 信号每帧中第  $n$  个脉冲的序列值，借助扩频序列的作用，不仅可使 UWB 系统的抗干扰能力进一步提高，还可实现 UWB 的多址通信； $T_f$  表示信息数据的周期，即一个信息符号所持续的时间长度； $T_c$  表示脉冲重复周期，也可理解为所采用扩频码的一个 chirp 宽度，且  $T_f = N_s \times T_c$ ，即每个信息符号用  $N_s$  个脉冲表示。其他参数的意义与式 8-4 中的介绍相同。采用一阶高斯脉冲作为 UWB 系统的发射脉冲波形的 DS-UWB 波形如图 8-4 所示。

图 8-4 中的脉冲位置是等间隔的，一般在 DS-UWB 的经典设计中，采用 1 个信息符号对应 1 个伪码的完整周期，而近些年来，研究人员开始将 DS-UWB 结合 TH 跳时进行混合调制，试图将 UWB 系统的多址干扰进一步减弱，并使 UWB 系统的发射功率谱更加的平滑，从而减少对共存通信系统的干扰。有关这方面的研究，读者可参考相关文献进行进一步的了解和学习。

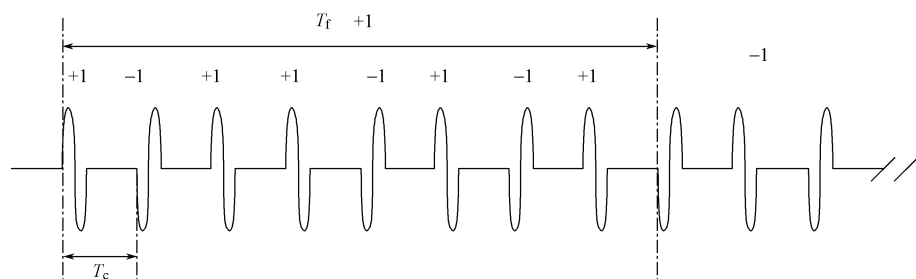


图 8-4 DS-UWB 系统发射脉冲波形示意图（采用一阶高斯脉冲）

## 8.3 UWB 关键技术介绍

### 8.3.1 脉冲信号的产生技术

从本质上讲，产生脉冲宽度为纳秒级（ $10^{-9}$  s）的信号源是 UWB 技术的前提条件，单个无载波窄脉冲信号有两个特点：一是激励信号的波形为具有陡峭前后沿的单个短脉冲，二是激励信号包括从直流到微波的很宽的频谱。

目前产生脉冲源的方法有：光电方法和电子方法。

光电方法的基本原理是利用光导开关的陡峭上升/下降沿获得脉冲信号。由激光脉冲信号激发得到的脉冲宽度可达到皮秒（ $10^{-12}$  s）量级，是最有发展前景的一种方法。

电子方法的基本原理是利用晶体管 PN 结反向加电，在雪崩状态的导通瞬间获得陡峭上升沿，整形后获得极短脉冲，是目前应用最广泛的方案。受晶体管耐压特性的限制，这种方法一般只能产生几十伏到上百伏的脉冲，脉冲的宽度可以达到 1 ns 以下，实际通信中使用一长串的超短脉冲。

### 8.3.2 UWB 的调制技术

由于 UWB 的传输功率受传输信号的功率谱密度限制，因而在两个方面影响调制方式的选择：一是对于每比特能量调制需要提供最佳的误码性能；二是影响了信号功率谱密度的结构，有可能把一些额外的限制加在传输功率上。在 UWB 中，信息是调制在脉冲上传递的，既可以用单个脉冲传递不同的信息，也可以使用多个脉冲传递相同的信息。

#### 1. 单脉冲调制

对于单个脉冲，脉冲的幅度、位置和极性变化都可以用于传递信息。适用于 UWB 的主要单脉冲调制技术有：脉冲幅度调制（PAM）、脉冲位置调制（PPM）、通断键控（OOK）、二相调制（BPM）和跳时/直扩二进制相移键控调制 TH/DS-BPSK 等。

PAM 是通过改变脉冲幅度的大小来传递信息的一种脉冲调制技术。PAM 既可以改变脉冲幅度的极性，也可以仅改变脉冲幅度的绝对值大小。通常所讲的 PAM 只改变脉冲幅度的绝对值。

BPM 和 OOK 是 PAM 的两种简化形式。BPM 通过改变脉冲的正负极性来调制二元信

息，所有脉冲幅度的绝对值相同。OOK 通过脉冲的有无来传递信息。在 PAM、BPM 和 OOK 调制中，发射脉冲的时间间隔是固定不变的。

实际上，我们也可以通过改变发射脉冲的时间间隔或发射脉冲相对于基准时间的位置来传递信息，这就是 PPM 的基本原理。在 PPM 中，脉冲的极性和幅度都不改变。

PAM、OOK 和 PPM 共同的优点是可以通过非相干检测恢复信息。PAM 和 PPM 还可以通过多个幅度调制或多个位置调制提高信息传输速率。然而 PAM、OOK 和 PPM 都有一个共同的缺点：经过这些方式调制的脉冲信号将出现线谱。线谱不仅会使 UWB 脉冲系统的信号难以满足一定的频谱要求（例如 FCC 关于 UWB 信号频谱的规定），而且还会降低功率的利用率。

通过上面 5 种调制方式的分析及实践中的应用可知：对于功率谱密度受约束和功率受限的 UWB 脉冲无线系统，为了获得更好的通信质量或更高的通信容量，BPM 是一种较理想的脉冲调制技术。

## 2. 多脉冲调制

在实际使用中，我们常使用多脉冲来提高抗干扰性能。当采用多脉冲调制时，传输相同信息的多个脉冲称为一组脉冲，多脉冲调制过程可以分两步：

第一步，每组脉冲内部的单个脉冲通常采用 PPM 或 BPM 调制；

第二步，每组脉冲作为整体通常可以采用 PAM、PPM 或 BPM 调制。

一般把第一步称为扩谱，而把第二步称为信息调制。因而在第一步中，把 PPM 称为跳时扩谱（TH-SS），即每组脉冲内部的每一个脉冲具有相同的幅度和极性，但具有不同的时间位置；把 BPM 称为直接序列扩谱（DS-SS），即每组脉冲内部的每一个脉冲具有固定的时间间隔和相同的幅度，但具有不同的极性。在第二步中，根据需要传输的信息比特，PAM 同时改变每组脉冲的幅度，PPM 同时调节每组脉冲的时间位置，BPM 同时改变每组脉冲的极性。这样，把第一步和第二步组合起来不难得到以下多脉冲调制技术：TH-SS PPM、DS-SS PPM、TH-SS PAM、DS-SS PAM、TH-SS BPM 和 DS-SS BPM 等。多脉冲调制不仅可以通过提高脉冲重复频率来降低单个脉冲的幅度或发射功率，更重要的是，多脉冲调制可以利用不同用户使用的 SS 序列之间的正交性或准正交性实现多用户干扰抑制，也可以利用 SS 序列的伪随机性实现窄带干扰抑制。在多脉冲调制中，利用不同 SS 序列之间的正交性，还可以通过同时传输多路多脉冲调制的信号来提高系统的通信速率，这样的技术通常被称为码分复用（CDMA）技术。2004 年的国际信号处理会议上提出了一种特殊的 CDMA 系统——无载波的正交频分复用系统（CL-UWB/OFDM），这种多脉冲调制技术可以有效地抑制多路数据之间的干扰和窄带干扰。

### 8.3.3 UWB 多址技术

在 UWB 系统中，多址接入方式与调制方式有密切联系。当系统采用 PPM 调制方式时，多址接入方式多采用跳时多址；若系统采用 BPSK 方式，多址接入方式通常有直序和跳时两种方式。基于上述两种基本的多址方式，许多其他多址方式陆续被提出，主要包括以下几种。

### 1. 伪混沌跳时多址方式 (PCTH)

PCTH 根据调制的数据,产生非周期的混沌编码,用它替代 TH-PPM 中的伪随机序列和调制的数据,控制短脉冲的发送时刻,使信号的频谱发生变化。PCTH 调制不仅能减少对现有无线通信系统的影响,而且更不易被检测到。

### 2. DS-BPSK/TH 混合多址方式

此方式在跳时 (TH) 的基础之上,通过直接序列扩频码进一步减少多址干扰,其多址性能优于 TH-PPM,与 DS-BPSK 相当,但在实现同步和抗远近效应方面,具有一定的优势。

### 3. DS-BPSK/Fixed TH 混合多址方式

此方式的特点是打破 TH-PPM 多址方式中采用随机跳时码的常规思路,利用具有特殊结构的固定跳时码,减少不同用户脉冲信号的碰撞概率。即使有碰撞发生时,利用直接序列扩频的伪随机码的特性,也可以进一步削弱多址干扰。

此外,由于 UWB 脉冲信号具有极低的占空比,其频谱能够达到 GHz 的数量级,因而 UWB 在时域中具有其他调制方式所不具有的特性。当多个用户的 UWB 信号被设计成具有不同的正交波形时,根据多个 UWB 用户时域发送波形的正交性来区分用户,实现多址,这被称之为波分多址技术。

## 8.4 UWB 应用介绍

由于 UWB 通信利用了一个相当宽的带宽,就好像使用了整个频谱,并且它能够与其他的应用共存,因此 UWB 可以应用在很多领域,如个域网、智能交通系统、无线传感网、射频识别、成像应用等。

### 1. UWB 在个域网中的应用

UWB 可以在限定的范围内 (如 4 m) 以很高的数据速率 (如 480 Mb/s)、很低的功率 (200  $\mu$ W) 传输信息,这比蓝牙好很多。蓝牙的数据速率是 1 Mb/s,功率是 1 mW。UWB 能够提供快速的无线外设访问来传输照片、文件、视频。因此 UWB 特别适合于个域网。通过 UWB,可以在家里和办公室里方便地以无线的方式将视频摄像机中的内容下载到 PC 中进行编辑,然后送到 TV 中浏览,轻松地以无线的方式实现掌上电脑 (PDA)、手机与 PC 数据同步,装载游戏和音频/视频文件到 PDA,音频文件在 MP3 播放器与多媒体 PC 之间传送,等等,如图 8-5 所示。

### 2. UWB 在智能交通信息中的应用

由于 UWB 技术具有一次突发大于 100 Mb/s 的数据速率,利用 UWB 可还可以建立智能交通管理系统,这种系统应该由若干个站台装置和一些车载装置组成无线通信网,两种装置之间通过 UWB 进行通信完成各种功能。



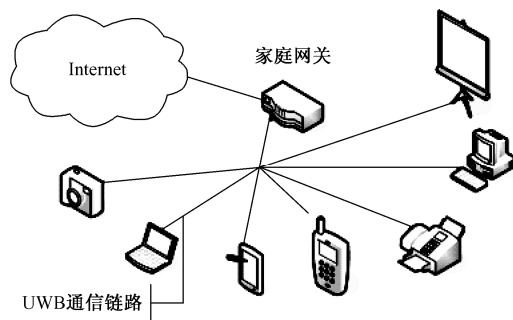


图 8-5 利用 UWB 技术构造的智能家庭网络示意图

例如，将公路上的信息（如路况、建筑物、天气预报等信息）发给路过汽车内的乘客，从而使行车更加安全、方便，也可实现不停车的自动收费、对汽车的定位搜索和速度测量等。

利用 UWB 的定位和搜索能力，可以制造防碰和防障碍物的雷达。装载了这种雷达的汽车会非常容易驾驶。当汽车的前方、后方、旁边有障碍物时，该雷达会提醒司机。在停车的时候，这种基于 UWB 的雷达是司机强有力的助手，如图 8-6 所示。

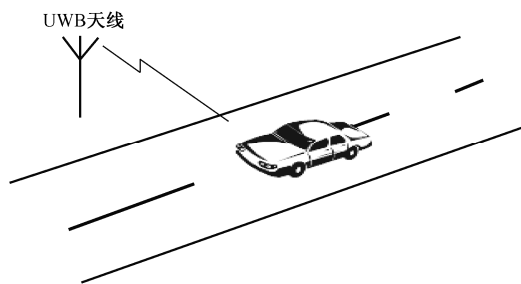


图 8-6 UWB 公路信息服务系统示意图

### 3. 传感器联网

利用 UWB 低成本、低功耗的特点，可以将 UWB 用于无线传感网。在大多数的应用中，传感器被用在特定的局域场所。传感器通过无线的方式而不是有线的方式传输数据将特别方便。作为无线传感网的通信技术，它必须是低成本的；同时它应该是低功耗的，以免频繁地更换电池。UWB 是无线传感网通信技术的最合适候选者。

### 4. 成像应用

由于 UWB 具有好的穿透墙、楼层的能力，UWB 可以应用于成像系统。利用 UWB 技术，可以制造穿墙雷达、穿地雷达。穿墙雷达可以用在战场上和警察的防暴行动中，定位墙后和角落的敌人；地面穿透雷达可以用来探测矿产，在地震或其他灾难后搜寻幸存者。基于 UWB 的成像系统也可以用于避免使用 X 射线的医学系统。

### 5. 军事应用

在军事方面，UWB 可用来实现战术/战略无线多跳网络电台，服务于战场自组织网络通信；也可用来实现非视距 UWB 电台，完成海军舰艇通信；还可以用于飞机内部通信，如有

效取代电缆的头盔。图 8-7 为空中防撞预警系统以及空中飞行器与地面的 UWB 数据传输示意图。

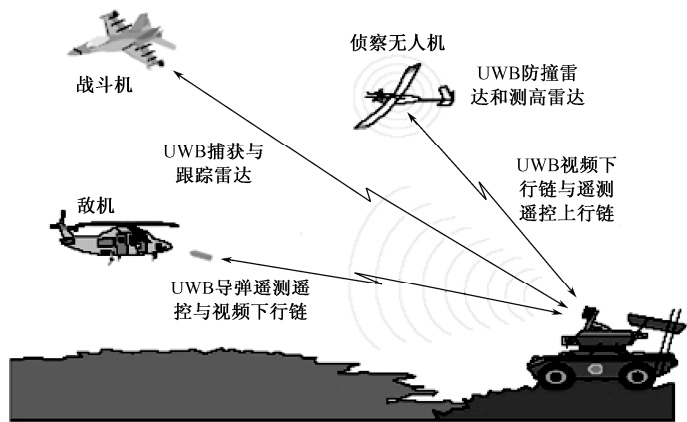


图 8-7 空中防撞预警系统以及空中飞行器与地面的 UWB 数据传输示意图

思考与练习

- 8-1 试说明 FCC 发布的超宽带信号的定义。
- 8-2 试说明超宽带通信的特点。
- 8-3 画图说明 UWB 发射机与接收机系统框图。
- 8-4 写出基于跳时体制的脉冲位置调制信号的数学表达式，说明其各个符号的意义，并给出时域波形图。

## 第9章 差错控制与信道编码技术

信道编码又称差错控制编码、可靠性编码、抗干扰编码或纠错码，它是提高数字信号传输可靠性的有效方法之一。

在无线数据传输过程中，引起数字传输差错的根本原因在于信道内存在的噪声，以及信道传输特性不理想所造成的码间串扰，为了提高数字传输系统的可靠性，降低信息传输的差错率，我们可以利用均衡技术消除码间串扰，利用增大发射功率、降低接收设备本身的噪声、选择好的调制制度和解调方法、加强天线的方向性等措施，提高数字传输系统的抗噪性能，但上述措施受到各种条件制约，只能将传输差错减小到一定程度。若希望进一步提高数字传输系统的可靠性，就可以采用差错控制编码，对可能或已经出现的差错进行控制。

### 9.1 差错控制编码的基本概念

#### 9.1.1 差错控制编码与信源编码的区别

为了提高数据信号传输的有效性而采取的编码称为信源编码，它通过科学合理的对信源进行代码分配，以尽量减少数据传输长度，提高编码效率为目的。

差错控制编码是为了提高数字通信的可靠性而采取的编码。一般在信息序列上，按照一定的规则加入若干监督码元，作为信息冗余，这些加入的码元与原来的信息码序列之间存在着某种确定的约束关系，使原来不规律的或规律性不强的原始数字信号，变为有规律的数字信号。在接收数据时，检验信息码元与监督码元之间的既定的约束关系，如该关系遭到破坏，则接收端可以发现传输中的错误，乃至纠正错误。

一般在  $k$  位信息码后面加  $r$  位监督码构成一个码组，码组的码位数为  $n$ ，即

$$\text{信息码}(k) + \text{监督码}(r) = \text{码组}(n) \quad (9-1)$$

可以看出，差错控制的方法来是以牺牲有效性为代价换取的。

#### 9.1.2 错误分类与错误图样

在无线数据传输过程中，按照错误的实际效果可分为单个码元的随机错误和多个码元的突发错误两种。

##### 1. 随机错误

随机错误又称独立差错，它是指那些独立的、稀疏的和互不相关的错误发生。随机噪声是导致出现随机错误的主要因素。存在这种错误的信道称为无记忆信道或随机信道，如微波接力信道和卫星转发信道等。

##### 2. 突发错误

突发错误是指一串串，甚至是成片出现的错误，错误之间有相关性，错误出现是密集的。

例如，传输的数据序列为 1111111...，由于噪声干扰，接收端收到的数据序列为 1001011...，其中 0010 为一串互相关联的错误，即一个突发错误。突发长度即为第一个错误与最后一个错误之间的长度（中间可能有少数不错的码），本例中突发长度等于 4。脉冲噪声是导致出现突发错误的主要因素。产生突发错误的信道称为有记忆信道或突发信道，例如短波、散射信道等。突发错误也可能是由信息存储载体缺陷（坏道）或电路接触不良等所引起的。

实际信道是复杂的，所出现的错误也不是单一的，而是随机和突发错误并存的，只不过有的信道以某种错误为主而已，这两类错误形式并存的信道称为组合信道或复合信道。

3. 错误图样

错误图样也称错误矢量，是发送的数据序列与接收序列对应码位的模 2 和，可直观地表示错误码元位置，即

错误图样  $E = \text{发送码组 } A \oplus \text{接收码组 } B$  (9-2a)

或

接收码组  $B = \text{发送码组 } A \oplus \text{错误图样 } E$  (9-2b)

例如：设发送码组  $A_1 = [a_6, a_5, a_4, a_3, a_2, a_1, a_0] = [1, 1, 1, 1, 1, 1, 1]$   
接收码组  $B_1 = [b_6, b_5, b_4, b_3, b_2, b_1, b_0] = [1, 1, 1, 1, 0, 1, 1]$   
错误图样  $E_1 = [e_6, e_5, e_4, e_3, e_2, e_1, e_0] = [0, 0, 0, 0, 1, 0, 0]$   
设发送码组  $A_2 = [a_6, a_5, a_4, a_3, a_2, a_1, a_0] = [1, 1, 1, 1, 1, 1, 1]$   
接收码组  $B_2 = [b_6, b_5, b_4, b_3, b_2, b_1, b_0] = [1, 0, 0, 1, 0, 1, 1]$   
错误图样  $E_2 = [e_6, e_5, e_4, e_3, e_2, e_1, e_0] = [0, 1, 1, 0, 1, 0, 0]$

显然，错误图样中“0”表示这位码无错，“1”表示有错。其中， $E_1$  为一个典型随机差错图样， $E_2$  为一个典型长度 4 的突发差错图样。

9.1.3 差错控制方式

常用的差错控制方式主要有 4 种：检错重发（ARQ）、前向纠错（FEC）、混合纠错（HEC）和信息反馈（IRQ），它们的结构如图 9-1 所示。

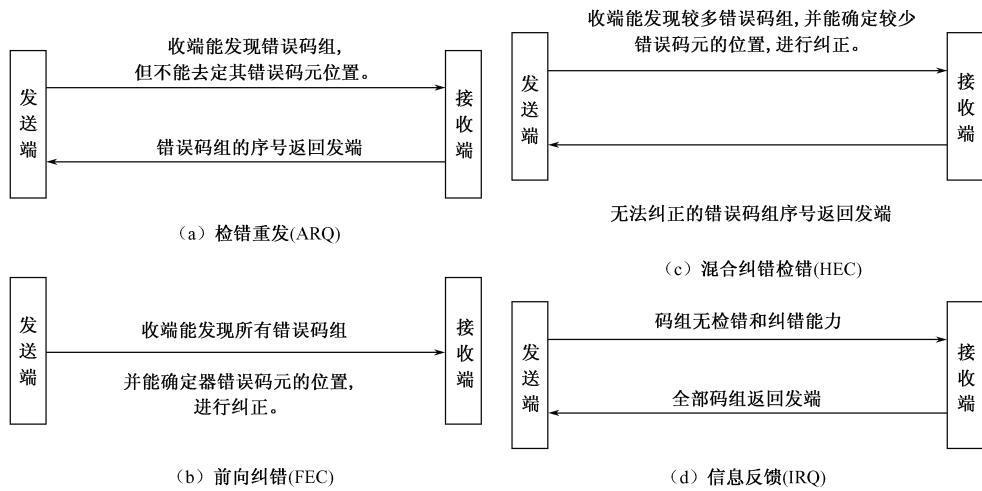


图 9-1 差错控制方式

# 1. 检错重发（ARQ）

在检错重发方式中，发送端经信道编码后可以发出能够检测出错误能力的码组；接收端收到后经检测如果发现传输中有错误，则通过反馈信道把这一判断结果反馈给发送端。然后，发送端把前面发出的信息重新传送一次，直到接收端认为已经正确后为止。典型系统原理方框图如图 9-2 所示，常用的检错重发系统有 3 种，即停发等候重发、返回重发和选择重发。

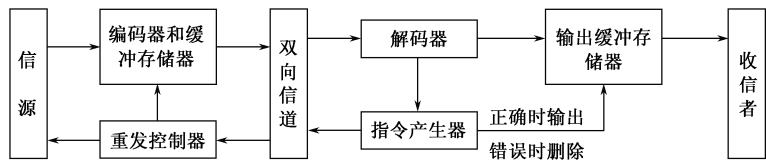


图 9-2 ARQ 系统组成方框图

停发等候重发系统的发送端在某一时刻向接收端发送一个码组，接收端收到后经检测若未发现传输错误，则发送一个认可信号（ACK）给发送端，发送端收到 ACK 信号后再发下一个码组；如果接收端检测出错误，则发送一个否认信号（NAK），发送端收到 NAK 信号后重发前一个码组，并再次等待 ACK 和 NAK 信号。这种方式效率不高，但工作方式简单，在计算机数据通信中仍在使用。

在返回重发系统中，发送端无停顿地送出一个又一个码组，不再等待 ACK 信号，一旦接收端发现错误并发回 NAK 信号，则发送端从下一个码组开始重发前一段  $N$  组信号， $N$  的大小取决于信号传递及处理所带来的延迟，这种系统比停发等候重发系统有很大的改进，在许多数据传输系统中得到应用。

在选择重发系统中，发送端也是连续不断地发送码组，接收端发现错误发回 NAK 信号。与返回重发系统不同的是，发送端不是重发前面的所有码组，而是只重发有错误的那一组。显然，这种选择重发系统传输效率最高，但控制最为复杂。此外，返回重发系统和选择重发系统都需要全双工的链路，而停发等候重发系统只需要半双工的链路。

根据不同思路，ARQ 还可以有其他的工作形式，如混合发送形式，它是将等候发送与连续发送结合起来的一种形式。发送端连续发送多个码组以后，再等待收端的应答信号，以决定是重发还是发送新的码组。在 ISO(国际标准化组织)建议的高级数据链路控制规程(HDLC)和 CCITT X.25 建议中就推荐采用这一工作形式。

基于上述分析，ARQ 的优点主要表现在：

- 只需少量的冗余码，就可以得到极低的输出误码率；
- 使用的检错码基本上与信道的统计特性无关，有一定的自适应能力；
- 与 FEC 相比，其信道编译码器的复杂性要低得多。

同时它也存在某些不足，主要表现在：

- 需要反向信道，故不能用于单向传输系统，并且实现重发控制比较复杂；
- 当信道干扰增大时，整个系统有可能处在重发循环当中，因而通信效率低；
- 不大适合于严格实时传输系统。

## 2. 前向纠错（FEC）

前向纠错系统中，发送端的信道编码器将输入数据序列变换成能够纠正错误的码，接收端的译码器根据编码规律检验出错误的位置并自动纠正。前向纠错方式不需要反馈信道，特别适合于只能提供单向信道的场合。由于能自动纠错，不要求检错重发，因而延时小，实时性好。其缺点是所选择的纠错码必须与信道的错误特性密切配合，否则很难达到降低错码率的要求；为了纠正较多的错码，译码设备复杂，而要求附加的监督码元也较多，传输效果就低。因此，过去单独使用这种控制方式的不很多，但随着编码理论和微电子技术的发展，译码设备成本下降，加之这种方式具有能实现单向通信和控制电路简单的优点，因而在实际应用中日益增多。

## 3. 混合纠错（HEC）

混合纠错方式是前向纠错方式和检错重发方式的结合。在这种系统中接收端不但具有纠正错误的能力，而且对超出纠错能力的错误有检测能力。遇到后一种情况时，系统可以通过反馈信道要求发送端重发一遍。

混合纠错检错方式在实时性和译码复杂性方面是前向纠错和检错重发方式的折中，因而近年来，在数据通信系统中采用较多。

## 4. 信息反馈（IRQ）

信息反馈方式中收端把收到的数据序列全部由反向信道送回发端，发端比较发送的数据序列与送回的数据序列，从而发现是否有错误，并把认为错误的数据序列的原数据再次传送，直到发端没有发现错误为止。

这种方式的优点是，不需要纠错、检错的编译器，设备简单。缺点是需要和前向信道相同的反向信道，实时性差。另外，发送端需要一定容量的存储器以存储发送码组，环路时延越大，数据传输速率越高，所需存储容量越大。因此 IRQ 方式仅适用于传输速率较低，数据信道差错率较低，且具有双向传输线路及控制简单的系统中。

上述差错控制方式应根据实际情况合理选用。除 IRQ 方式外，都要求发送端发送的数据序列具有检错或纠错能力。为此，必须对信息输出的数据以一定规则加入多余码元（纠错编码）。对于纠错编码的要求是加入的多余码元少而纠错能力却很高，而且实现方便，设备简单，成本低。下面将讨论一些常用的纠错检错码。

### 9.1.4 最小码距与检（纠）错能力

为了说明检错与纠错的基本原理，以及最小码距与检（纠）错的关系，下面我们将首先介绍几个有关的概念。

码长：码组中码元的数目。

码重：码组中非 0 位的数目；对于二进制码来讲，码重  $W$  就是码元中 1 的数目，例如码组 10100，码长  $n=5$ ，码重  $W=2$ 。

码距：两个等长码组之间对应位不同的数目，称为这两个码组的距离，并将二维空间距离叫汉明距离，三维、四维等空间的几何距离叫欧氏距离。例如，码组 10100 与 11000 之间的汉明码距为  $d=2$ 。

传统的纠错编码是以二进制的汉明距离作为最佳的译码判决准则的。在对于非二进制信号时，并不能用汉明距离来表示该码的纠错能力，三维、四维等空间距离才是衡量性能的有效方法。

最小码距：在码组集合中全体码组之间距离的最小数值。

对于二进制码组而言，两个码组之间的模 2 相加，其不同的对应位必为 1，相同的对应位必为 0，因此，两个码组之间模二相加得到的信码组的重量就是这两个码组之间的汉明距离。码组之间的最小距离是衡量该码组检错和纠错能力的重要依据，因此，最小码距是信道编码的一个重要的参数。在一般情况下，对于分组码的最小汉明距离  $d_0$  与检错和纠错能力之间满足下列关系：

(1) 当码组用于检测错误时，如果要检测  $e$  个错误，则

$$d_0 \geq e + 1 \tag{9-3}$$

这个关系可以利用图 9-3 (a) 予以说明。其中用 A 和 B 分别表示两个码距为  $d_0$  的码组，若 A 发生  $e$  个错误，则 A 就变成以 A 为球心， $e$  为半径的球面上的码组，为了能将这组码分辨出来，它们必须距离其最近的码组 B 有一位的差别，即 A 和 B 之间最小距离为  $d_0 \geq e + 1$ 。

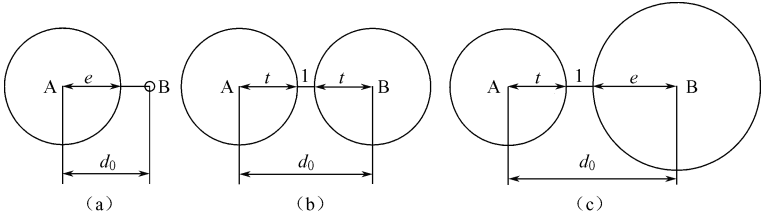


图 9-3 纠（检）错能力的几何解释

(2) 当码组用于纠正错误时，如果要纠正  $t$  个错误，则

$$d_0 \geq 2t + 1 \tag{9-4}$$

这个关系可以利用图 9-3 (b) 予以说明。在图中用 A 和 B 分别表示两个码距为  $d_0$  的码组，若 A 发生  $t$  个错误，则 A 就变成以 A 为球心， $t$  为半径的球面上的码组；B 发生  $t$  个错误，则 B 就变成以 B 为球心， $t$  为半径的球面上的码组。为了在出现  $t$  个错误之后，仍能够分辨出 A 和 B 来，那么，A 和 B 之间距离应大于  $2t$ ，最小距离也应当使两球体表面相距为 1，即满足不等式 (9-4)。

(3) 若码组用于纠  $t$  个错，同时检  $e$  个错时，则

$$d_0 \geq t + e + 1 \tag{9-5}$$

这个关系可以利用图 9-3 (c) 予以说明。在图中用 A 和 B 分别表示两个码距为  $d_0$  的码组，当码组出现  $t$  个或小于  $t$  个错时，系统按照纠错方式工作；当码组出现大于  $t$  个而小于  $e$  个错时，系统按照检错方式工作；若 A 发生  $t$  个错误，B 发生  $e$  个错误时，既要纠 A 的错，又要检 B 的错，则 A 和 B 之间距离应大于  $t + e$ ，也就是满足式 (9-5)。

通常，在信道编码过程中，监督位越多纠错能力就越强，但编码效率就越低。若码组中信息位数为  $k$ ，监督位数为  $r$ ，码长  $n = k + r$ ，则编码效率  $R_c$  可表示为：

$$R_c = k / n = (n - r) / n = 1 - r / n \tag{9-6}$$

信道编码的任务就是要根据不同的干扰特性，设计出编码效率高，纠错能力强的编码。在实际设计过程中，需要根据具体指标要求，尽量简化编码实际的复杂度，节省设计费用。

9.2 分组码

分组码是一组固定长度的码组，可表示为  $(n, k)$ ，通常它用于前向纠错。在分组码中，监督位被加到信息位之后，形成新的码。在编码时， $k$  个信息位被编为  $n$  位码组长度，而  $n-k$  个监督位的作用就是实现检错与纠错。

当分组码的码组采用二进制进行表示时，它就被称为二进制分组码。对于长度  $n$  的二进制分组码，它有  $2^n$  种可能的码组，从  $2^n$  种码组中，可以选择  $M=2^k$  个码组 ( $k < n$ ) 组成一种码。这样，一个  $k$  比特信息的分组码可以映射到一个长度为  $n$  码组上，该码组是从由  $M=2^k$  个码组构成的码集中选出来的，这样剩下的码组就可以对这个分组码进行检错或纠错，不同的信道编码方案所提供的差错控制能力也不相同。

9.2.1 线性分组码

当分组码的信息码元与监督码元之间的关系为线性关系时，这种分组码就被称为线性分组码。线性分组码是建立在代数群论基础之上的，各许用码的集合构成了代数学中的群，它们的主要性质如下：

- 任意两许用码之和（对于二进制码这个和的含义是模 2 和）仍为一许用码，也就是说，线性分组码具有封闭性；
- 码组间的最小码距等于非零码的最小码重。

下面通过一个例子说明线性分组码是如何构造的。

设分组码  $(n, k)$  中  $k=4$ ，为了能够纠正一位错误，要求  $r \geq 3$ ，取  $r=3$ ，则  $n=k+r=7$ 。因此，可以用  $a_6a_5a_4a_3a_2a_1a_0$  表示这 7 个码元，用  $S_3$ 、 $S_2$ 、 $S_1$  表示 3 个监督方程的计算得到的校正子，并假设  $S_3$ 、 $S_2$ 、 $S_1$  三位校正子码组与误码位置的关系如表 9-1 所示。

表 9-1 校正子与误码位置

| $S_1S_2S_3$ | 误码位置  | $S_1S_2S_3$ | 误码位置  |
|-------------|-------|-------------|-------|
| 001         | $a_0$ | 101         | $a_4$ |
| 010         | $a_1$ | 110         | $a_5$ |
| 100         | $a_2$ | 111         | $a_6$ |
| 011         | $a_3$ | 000         | 无错    |

根据表 9-1 的对应关系，可以得到下列逻辑关系式：

$$\begin{aligned} S_1 &= a_6 + a_5 + a_4 + a_2 \\ S_2 &= a_6 + a_5 + a_3 + a_1 \\ S_3 &= a_6 + a_4 + a_3 + a_0 \end{aligned}$$

(9-7)

在进行编码时，设  $a_6$ 、 $a_5$ 、 $a_4$ 、 $a_3$  为信息码元，从表 9-1 中可以看到，当  $S_3S_2S_1=000$  时，就表明码组在传输过程中没有发生错误，基于这一约束，利用式 (9-7) 的 3 个等式，可以得到下面两种形式的线性方程组：

$$\begin{cases} a_6 + a_5 + a_4 + a_2 = 0 \\ a_6 + a_5 + a_3 + a_1 = 0 \\ a_6 + a_4 + a_3 + a_0 = 0 \end{cases}$$

(9-8a)



$$\begin{cases} a_6 + a_5 + a_4 = a_2 \\ a_6 + a_5 + a_3 = a_1 \\ a_6 + a_4 + a_3 = a_0 \end{cases} \quad (9-8b)$$

根据上面两个线性关系式，可以得到 16 个许用码组，如表 9-2 所示。

表 9-2 许用码组

| 信息位            | 监督位         | 信息位            | 监督位         | 信息位            | 监督位         | 信息位            | 监督位         |
|----------------|-------------|----------------|-------------|----------------|-------------|----------------|-------------|
| $a_6a_5a_4a_3$ | $a_2a_1a_0$ | $a_6a_5a_4a_3$ | $a_2a_1a_0$ | $a_6a_5a_4a_3$ | $a_2a_1a_0$ | $a_6a_5a_4a_3$ | $a_2a_1a_0$ |
| 0000           | 000         | 0100           | 110         | 1000           | 111         | 1100           | 001         |
| 0001           | 011         | 0101           | 101         | 1001           | 100         | 1101           | 010         |
| 0010           | 101         | 0100           | 011         | 1010           | 010         | 1100           | 100         |
| 0011           | 110         | 0111           | 000         | 1011           | 001         | 1111           | 111         |

在接收端收到码组以后，就可以代入式 (9-7) 计算  $S_1$ 、 $S_2$  和  $S_3$ ，如果全为 0，则表明传输时没有发生错误，否则根据表 9-1 纠正错误。当然对于上述 (7, 4) 分组码而言，最小码距  $d_0=3$ ，因此，它可以纠正一个错误或检测两个错误，如果超出这个范围，纠错功能就要失败。

对于式 (9-8a) 可以用矩阵形式来表示

$$\begin{bmatrix} 1 & 1 & 1 & 0 & 1 & 0 & 0 \\ 1 & 1 & 0 & 1 & 0 & 1 & 0 \\ 1 & 0 & 1 & 1 & 0 & 0 & 1 \end{bmatrix} \cdot [a_6 \ a_5 \ a_4 \ a_3 \ a_2 \ a_1 \ a_0]^T = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix} \quad (9-9)$$

式 (9-9) 可以记为  $\mathbf{H}\mathbf{A}^T=\mathbf{0}^T$  或  $\mathbf{A}\mathbf{H}^T=\mathbf{0}$ ，其中

$$\mathbf{H} = \begin{bmatrix} 1 & 1 & 1 & 0 & 1 & 0 & 0 \\ 1 & 1 & 0 & 1 & 0 & 1 & 0 \\ 1 & 0 & 1 & 1 & 0 & 0 & 1 \end{bmatrix} = [\mathbf{P} \ \mathbf{I}_r] \quad (9-10)$$

$$\mathbf{A} = [a_6 \ a_5 \ a_4 \ a_3 \ a_2 \ a_1 \ a_0], \quad \mathbf{0} = [0 \ 0 \ 0]。$$

通常  $\mathbf{H}$  被称为监督矩阵， $\mathbf{A}$  被称为信道编码得到的码字。在这个例子中  $\mathbf{H}$  为  $r \times n$  阶矩阵， $\mathbf{P}$  为  $r \times k$  阶矩阵， $\mathbf{I}_r$  为  $r \times r$  阶单位矩阵，具有这种特性的  $\mathbf{H}$  矩阵被称为典型监督矩阵，这是一种较为简单的信道编译码方式。典型形式的监督矩阵各行是线性无关的，非典型形式的监督矩阵可以经过行或列的运算化为典型形式。

式 (9-8b) 可用矩阵形式来表示为：

$$\begin{bmatrix} a_1 \\ a_2 \\ a_3 \end{bmatrix} = \begin{bmatrix} 1 & 1 & 1 & 0 \\ 1 & 1 & 0 & 1 \\ 1 & 0 & 1 & 1 \end{bmatrix} \cdot \begin{bmatrix} a_6 \\ a_5 \\ a_4 \\ a_3 \end{bmatrix} \quad (9-11a)$$

或者

$$[a_2 \ a_1 \ a_0] = [a_6 \ a_5 \ a_4 \ a_3] \cdot \begin{bmatrix} 1 & 1 & 1 \\ 1 & 1 & 0 \\ 1 & 0 & 1 \\ 0 & 1 & 1 \end{bmatrix} = [a_6 \ a_5 \ a_4 \ a_3] \cdot \mathbf{Q} \quad (9-11b)$$

比较式 (9-9) 和式 (9-11b) 可以看到  $\mathbf{Q} = \mathbf{P}^T$ ，如果在  $\mathbf{Q}$  矩阵的左边再加上一个  $k \times k$

的单位矩阵，就形成了一个新矩阵  $G$ ：

$$G = [I_k \quad Q] = \begin{bmatrix} 1 & 0 & 0 & 0 & 1 & 1 & 1 \\ 0 & 1 & 0 & 0 & 1 & 1 & 0 \\ 0 & 0 & 1 & 0 & 1 & 0 & 1 \\ 0 & 0 & 0 & 1 & 0 & 1 & 1 \end{bmatrix} \quad (9-12)$$

$$A = M \cdot G = [a_6 \quad a_5 \quad a_4 \quad a_3] \cdot G \quad (9-13)$$

这里  $M$  为要发送的信息码元， $G$  称为生成矩阵，利用它可以产生整个码组。

由式 (9-12) 表示的生成矩阵形式被称为典型生成矩阵，利用式 (9-13) 产生的分组码为系统码，也就是信息码元保持不变，监督码元附加在其后。

在发送端信息码元  $M$  利用式 (9-13)，实现信道编码，产生线性分组码  $A$ ；在传输过程中有可能出现误码，设接收到的码组为  $B$ 。则收发码组之差为

$$B - A = [b_{n-1} \quad b_{n-2} \quad \cdots \quad b_0] - [a_{n-1} \quad a_{n-2} \quad \cdots \quad a_0] = E = [e_{n-1} \quad e_{n-2} \quad \cdots \quad e_0] \quad (9-14)$$

这里  $e_i = \begin{cases} 0 & b_i = a_i \\ 1 & b_i \neq a_i \end{cases}$ ， $e_i = 1$  表示  $i$  位有错， $e_i = 0$  表示  $i$  位无错。基于这样的原则接收端利用接收到的码组  $B$  计算校正子：

$$S = BH^T = (A + E)H^T = AH^T + EH^T = EH^T \quad (9-15)$$

因此，校正子仅与  $E$  有关，即错误图样与校正子之间有确定的关系。

在实践中经常会遇到的线性分组码主要有以下几种：

### 1. 汉明 (Hamming) 码

汉明码既有二进制的，也有非二进制的，这里仅讨论二进制汉明码的性质。二进制汉明码可以表示为

$$(n, k) = (2^m - 1, 2^m - 1 - m) \quad (9-16)$$

这里  $m$  可取大于等于 2 的任意整数，因此，汉明码的特点如下：码长  $n = 2^m - 1$ ，信息位  $k = 2^m - 1 - m$ ，监督位  $r = m$ ，最小码距  $d_0 = 3$ ，纠错能力  $t = 1$ 。如果要产生一个系统汉明码，可以将矩阵  $H$  转换成典型形式的监督矩阵，进一步利用  $Q = P^T$  的关系，得到相应的生成矩阵  $G$ 。

### 2. 哈达码 (Hadamard) 码

哈达码矩阵  $M_n$  是一个由“0”和“1”构成的  $n \times n$  维矩阵 ( $n$  是偶数)，矩阵中任意两行相比较，都存在  $n/2$  个不同的元素，矩阵中有一行是全 0 行，其他行都包含  $n/2$  个“0”和  $n/2$  个“1”。当  $n=2$  时，哈达码矩阵可表示为

$$M_2 = \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix} \quad (9-17)$$

进一步可以按照下述规律由  $M_n$  产生哈达码矩阵  $M_{2n}$

$$M_{2n} = \begin{bmatrix} M_n & M_n \\ M_n & \overline{M_n} \end{bmatrix} \quad (9-18)$$

式中， $\overline{M_n}$  为  $M_n$  的互补矩阵。按上述规律， $M_4$  和  $\overline{M_n}$  就可以分别表示为

$$\mathbf{M}_4 = \begin{bmatrix} \mathbf{M}_2 & \mathbf{M}_2 \\ \mathbf{M}_2 & \overline{\mathbf{M}_2} \end{bmatrix} = \begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 1 \\ 0 & 0 & 1 & 1 \\ 0 & 1 & 1 & 0 \end{bmatrix}, \quad \overline{\mathbf{M}_4} = \begin{bmatrix} 1 & 1 & 1 & 1 \\ 1 & 0 & 1 & 0 \\ 1 & 1 & 0 & 0 \\ 1 & 0 & 0 & 1 \end{bmatrix} \quad (9-19)$$

至此，可以利用  $\mathbf{M}_4$  和  $\overline{\mathbf{M}_4}$  的行构成一个码长  $n=4$  的二进制线性分组码，该码由 8 个码组构成，最小码距为  $d_0=2$ 。基于这种方法构成的码，就被称为哈达码码。因此，哈达码码的长度  $n=2^m$ ， $k = \log_2 2n = \log_2 2^{m+1} = m+1$ ， $d_0 = n/2 = 2^{m-1}$ ，这里  $m$  表示正整数。

## 9.2.2 循环码

循环码是线性码的一个重要子集，是目前研究得最成熟的一类码。它有许多特殊的代数性质，这些性质有助于按所要求的纠错能力，系统地构造这类码，且易于实现；同时循环码的性能也较好，具有较强的检错和纠错能力。

循环码最大的特点就是循环性，所谓循环性是指：循环码中任一许用码组经过循环移位后，所得到的码组仍然是许用码组。若  $(a_{n-1} a_{n-2} \cdots a_1 a_0)$  为一循环码组，则  $(a_{n-2} a_{n-3} \cdots a_0 a_{n-1})$ 、 $(a_{n-3} a_{n-4} \cdots a_{n-1} a_{n-2})$ 、 $\cdots$  还是许用码组。也就是说，不论是左移或是右移，也不论移多少位，其所的码组仍然是许用的循环码组。

为了利用代数理论研究循环码，可以将码组用代数多项式来表示，这个多项式被称为码多项式，对于许用循环码  $\mathbf{A} = (a_{n-1} a_{n-2} \cdots a_1 a_0)$ ，可以将它的码多项式表示为

$$A(x) = a_{n-1}x^{n-1} + a_{n-2}x^{n-2} + \cdots + a_1x + a_0 \quad (9-20)$$

对于二进制码组，多项式的每个系数不是 0 就是 1， $x$  仅是码元位置的标志，因此，我们这里并不关心  $x$  的取值。

设上述循环许用码组  $\mathbf{A}$  左循环一位得到的码组记为  $\mathbf{A}^{(1)} = (a_{n-2} a_{n-3} \cdots a_0 a_{n-1})$ ，其码多项式可以表示为

$$A^{(1)}(x) = a_{n-2}x^{n-1} + a_{n-3}x^{n-2} + \cdots + a_0x + a_{n-1} \quad (9-21)$$

同理，左移  $i$  位的码组  $\mathbf{A}^{(i)} = (a_{n-i-1} a_{n-i-2} \cdots a_{n-i+1} a_{n-i})$ ，其码多项式为

$$A^{(i)}(x) = a_{n-i-1}x^{n-1} + a_{n-i-2}x^{n-2} + \cdots + a_{n-i+1}x + a_{n-i} \quad (9-22)$$

利用代数理论知识， $A^{(i)}(x)$  也可以用下式求得：

$$x^i \cdot A(x) = Q(x) \cdot (x^n + 1) + A^{(i)}(x) \quad (9-23)$$

式中， $Q(x)$  表示  $x^i \cdot A(x)$  除以  $(x^n + 1)$  的商，而  $A^{(i)}(x)$  表示所得余式。由上述分析可以得到结论：一个长为  $n$  的循环码，它必为按模  $(x^n + 1)$  运算的一个余式。这个结论给出了构造许用码的一种方法，也就是利用循环码的生成多项式可以得到全部码组。

在循环码中，一个  $(n, k)$  码有  $2^k$  个不同的码组，若用  $g(x)$  表示其中前  $(k-1)$  位皆为“0”码组，则  $g(x)$ 、 $x \cdot g(x)$ 、 $x^2 \cdot g(x)$ 、 $\cdots$ 、 $x^{k-1} \cdot g(x)$  都是码组，而且这  $k$  个码组线性无关，因此，可以利用它们构成循环码的生成矩阵，而  $g(x)$  被称为生成多项式。

可以证明，一个  $(n, k)$  循环码的生成多项式  $g(x)$ ，必须是一个常数项不为“0”的  $(n-k)$  次多项式，而且，这个  $g(x)$  还是  $(n, k)$  码中次数为  $(n-k)$  的唯一的—一个多项式。因为，如果有两个，则由于码的封闭性，把这两个码相加也应该是一个码组，且此码组多项式的次数将小于  $(n-k)$ ，即出现连续“0”的个数将多于  $(k-1)$  的情况，这与  $(n, k)$  循环码是

线性码的特性相违背，故是不可能的。为此，可以得到一个重要的结论：一旦生成多项式  $g(x)$  确定以后，整个  $(n, k)$  循环码就被确定了。基于  $g(x)$  可以进一步写出循环码的生成矩阵如下：

$$G(x)=\begin{bmatrix} x^{k-1}\cdot g(x) \\ x^{k-2}\cdot g(x) \\ \vdots \\ x\cdot g(x) \\ g(x) \end{bmatrix} \tag{9-24}$$

显然，式（9-24）不符合  $G=[I_k \quad Q]$  形式，所以此生成矩阵不是典型形式，不过，可以通过简单的代数变换将它变成典型矩阵。

根据  $g(x)$  定义可知，它是  $(n, k)$  循环码中唯一的一个  $(n-k)$  次码多项式，当然也就是该循环码的许用码组。对  $g(x)$  表述的许用码组进行  $k$  次左移，会得到同样的码组，将这一过程用式（9-23）描述将得到

$$x^k \cdot g(x) = Q(x) \cdot (x^n + 1) + g(x) \tag{9-25}$$

式（9-25）左边是一个  $n$  次多项式，因此  $Q(x)=1$ ，所以可以进一步表示为

$$(x^n + 1) = x^k \cdot g(x) + g(x) = g(x) \cdot (x^k + 1) \tag{9-26}$$

式（9-26）表明，生成多项式  $g(x)$  是  $(x^n+1)$  的一个因式，因此，为了确定生成多项式，必须首先对  $(x^n+1)$  进行因式分解，然后通过计算进行筛选，计算过程通常使用计算机来完成。

对于一个  $(n, k)$  循环码，其生成多项式应该是  $(x^n+1)$  的一个  $(n-k)$  次因子，任何  $(n, k)$  循环码的生成多项式  $g(x)$ ，乘上  $(x+1)$  后得到生成多项式，可以构造  $(n, k-1)$  循环码。以  $(x^7+1)$  因式分解为例，

$$x^7 + 1 = (x + 1) \cdot (x^3 + x + 1) \cdot (x^3 + x^2 + 1) \tag{9-27}$$

由式（9-27）可构成如表 9-3 所示的  $(7, k)$  循环码。

表 9-3  $(x^7+1)$  因式分解构成的  $(7, k)$  循环码

| $(n, k)$ | $g(x)$                                | $h(x)$                                |
|----------|---------------------------------------|---------------------------------------|
| (7,6)    | $x+1$                                 | $(x^3+x^2+1)(x^3+x+1)$                |
| (7,4)    | $x^3+x^2+1$ 或 $x^3+x+1$               | $(x^3+x+1)(x+1)$ 或 $(x^3+x^2+1)(x+1)$ |
| (7,3)    | $(x^3+x^2+1)(x+1)$ 或 $(x^3+x+1)(x+1)$ | $x^3+x+1$ 或 $x^3+x^2+1$               |
| (7,1)    | $(x^3+x^2+1)(x^3+x+1)$                | $x+1$                                 |

可以看出，不管是  $(7, 4)$  循环码还是  $(7, 3)$  循环码均包含两个不同的生成多项式，因此，依据不同的生成多项式将产生不同的循环码组。

上面我们讨论了循环码的基本原理，下面就系统循环码的产生进行分析。

根据循环码的编码特点，所有循环码多项式  $A(x)$  都可以被  $g(x)$  整除。根据这一原理可以得到一个较简单的系统循环码编码方法：设要产生  $(n, k)$  循环码， $m(x)$  表示信息多项式，则其次数必小于  $k$ ，而  $x^{n-k} \cdot m(x)$  的次数必小于  $n$ ，用  $x^{n-k} \cdot m(x)$  除以  $g(x)$ ，可得余数  $r(x)$ ， $r(x)$  的次数必小于  $(n-k)$ ，将  $r(x)$  加到信息位后作监督位，就得到了系统循环码，其数学描述如下：

$$\frac{x^{n-k} \cdot m(x)}{g(x)} = q(x) + \frac{r(x)}{g(x)} \tag{9-28}$$

则系统循环码可以表示成为

$$A(x) = x^{n-k} \cdot m(x) + r(x) \tag{9-29}$$

上述编码过程，在硬件实现时，可以利用除法电路来实现，这里的除法电路采用一些移位寄存器和模 2 加法器来构成。下面我们将以（7，3）循环码为例，来说明其具体实现过程。设该（7，3）循环码的生成多项式为： $g(x)=x^4+x^2+x+1$ ，则构成的系统循环码编码器如图 9-3 所示，图 9-3 中有 4 个移位寄存器，一个双刀双掷开关。当信息位输入时，开关位置接“1”，输入的信息码一方面送到除法器进行运算，一方面直接输出；当信息位全部输出后，开关位置接“2”，这时输出端接到移位寄存器的输出，这时除法的余项，也就是监督位依次输出。当信息码为 110 时，编码器的工作过程如表 9-4 所示。

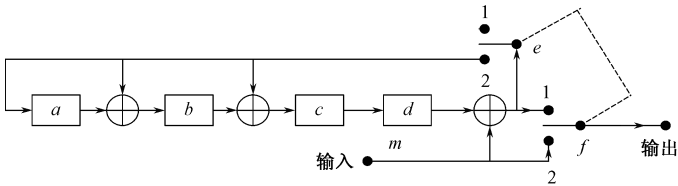


图 9-4 （7，3）循环码编码器

表 9-4 编码器的工作过程

| 输入 (m) | 移位寄存器 (abcd) | 反馈 (e) | 输出 (f) |
|--------|--------------|--------|--------|
| 0      | 0000         | 0      | 0      |
| 1      | 1110         | 1      | 1      |
| 1      | 1001         | 1      | 1      |
| 0      | 1010         | 1      | 0      |
| 0      | 0101         | 0      | 0      |
| 0      | 0010         | 1      | 1      |
| 0      | 0001         | 0      | 0      |
| 0      | 0000         | 1      | 1      |

对于接收端译码的要求通常有两个：检错与纠错。达到检错目的的译码十分简单，可以依据式（9-25），通过判断接收到的码组多项式  $B(x)$  是否能被生成多项式  $g(x)$  整除作为依据。当传输中未发生错误时，也就是接收的码组与发送的码组相同，即  $A(x)=B(x)$ ，故接收的码组  $B(x)$  必能被  $g(x)$  整除；若传输中发生了错误，则  $A(x) \neq B(x)$ ， $B(x)$  不能被  $g(x)$  整除。因此，我们就可以根据余项是否为零来判断码组中是否有错码。

需要指出的是，有错码的接收码组也有可能被  $g(x)$  整除，这时的错码就不能检出了。这种错误被称为不可检错误，不可检错误的错码数必将超过这种编码的检错能力。

在接收端为纠错而采用的译码方法自然比检错要复杂许多，因此，对纠错码的研究大都集中在译码算法上。我们知道，校正子与错误图样之间存在某种对应关系。如同其他线性分组码，循环码的译码可以分 3 步进行：

- （1）由接收到的码多项式  $B(x)$  计算校正子（伴随式）多项式  $S(x)$ ；
- （2）由校正子  $S(x)$  确定错误图样  $E(x)$ ；

(3) 将错误图样  $E(x)$  与  $B(x)$  相加, 纠正错误。

上述第 (1) 步运算和检错译码类似, 也就是求解  $B(x)$  整除  $g(x)$  的余式, 第 (3) 步也很简单。因此, 纠错码译码器的复杂性主要取决于译码过程的第 (2) 步。

基于错误图样识别的译码器称为梅吉特译码器, 它的原理如图 9-5 所示。错误图样识别器是一个具有  $(n-k)$  个输入端的逻辑电路, 原则上可以采用查表的方法, 根据校正子找到错误图样, 利用循环码的上述特性可以简化识别电路。梅吉特译码器特别适合于纠正 2 个以下的随机独立错误。

图 9-5 中  $k$  级缓存器用于存储系统循环码的信息码元, 模 2 加电路用于纠正错误。当校正子为 0 时, 模 2 加来自错误图样识别电路的输入端为 0, 输出缓存器的内容; 当校正子不为 0 时, 模 2 加来自错误图样识别电路的输入端在第  $i$  位输出为 1, 它可以使缓存器输出取补, 即纠正错误。

循环码的译码方法除了梅吉特译码以外, 还有捕错译码、大数逻辑译码等方法。

捕错译码是梅吉特译码的一种变形, 也可以用较简单的组合逻辑电路实现, 它特别适合于纠正突发错误、单个随机错误和两个错误的码。

大数逻辑译码有时也被称为门限译码, 这种译码方法也很简单, 但它只能用于有一定结构的为数不多的码型, 虽然在一般情形下, 大数逻辑可译码的纠错能力和编码效率比有相同参数的其他循环码 (如 BCH 码) 稍差, 但它的译码算法和硬件比较简单, 因此在实际中有较广泛的应用。

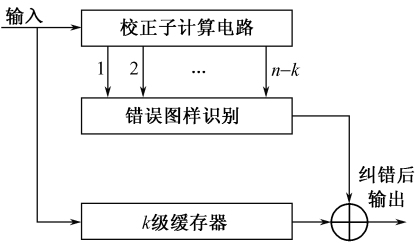


图 9-5 循环码梅吉特译码原理框图

### 9.3 卷积码

分组码  $(n, k)$  中包含  $k$  个信息位, 码组长度为  $n$ , 每个码组的  $(n-k)$  个校验位仅与本码组的  $k$  个信息位有关, 而与其他码组无关。为了达到一定的纠错能力和编码效率  $(R_c=k/n)$ , 分组码的码组长度  $n$  通常都比较大。编译码时必须把整个信息码组存储起来, 由此产生的延时随着  $n$  的增加而线性增加。

为了减少这个延迟, 人们提出了各种解决方案, 其中卷积码就是其中一种较好的信道编码方式。这种编码方式同样也是把  $k$  个信息比特编成  $n$  个比特, 但  $k$  和  $n$  通常很小, 特别适宜于以串行形式传输信息, 减小了编码延时。与分组码不同, 卷积码中编码后的  $n$  个码元不仅与当前段的  $k$  个信息有关, 而且也与前面  $(N-1)$  段的信息有关, 编码过程中相互关联的码元为  $nN$  个。卷积码的纠错能力随着  $N$  的增加而增大, 在编码器复杂程度相同的情况下, 卷积码的性能优于分组码。另一点不同的是: 分组码有严格的代数结构, 但卷积码至今尚未找到如此严密的数学手段, 把纠错性能与码的结构十分有规律地联系起来, 目前大都采用计算机来搜索好码。

#### 9.3.1 卷积码编码器及表述方法

卷积码编码器的一般形式如图 9-6 所示, 它包括: 一个由  $N$  段组成的输入移位寄存器,

每段有  $k$  级，共  $Nk$  位寄存器；一组  $n$  个模 2 相加器，一个由  $n$  级组成的输出移位寄存器。对应于每段 1 个比特的输入序列，输出  $n$  个比特。由图可知， $n$  个比特编码输出不仅与当前的  $k$  个比特信息输入有关，而且与以前的  $(N-1)k$  个比特信息输入有关。整个编码过程可以看成是输入信息序列与信道编码器的卷积，卷积码即由此得名。通常把  $N$  称为约束长度（注意：约束长度的定义并无统一的标准，在有的书和文献中把  $nN$  或  $(N-1)$  称为约束长度），因此，卷积码通常可以表示为  $(n, k, N)$ ，它的编码效率为  $R_c=k/n$ 。

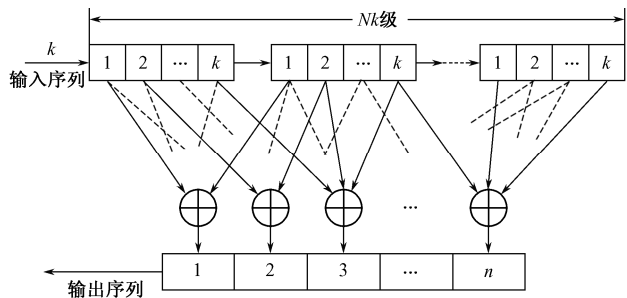


图 9-6 卷积码编码器一般形式

为了介绍几种简单的卷积码表述方法，我们将以图 9-7 所示的  $(2, 1, 3)$  卷积码为例进行分析。

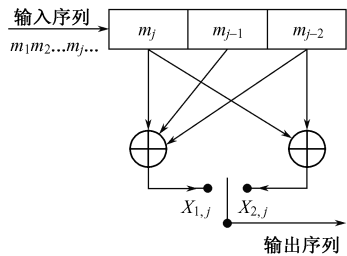


图 9-7  $(2, 1, 3)$  卷积码编码器

### 1. 树状图

在  $(2, 1, 3)$  卷积码编码器当中，输出移位寄存器用转换开关代替，每输入一个位信息经编码产生二位输出。假设移位寄存器的起始状态为全 0，当第一个输入位为 0 时，输出为 00；若输入比特为 1，则输出比特为 11。随着第二个位的输入，第一位右移一位，此时输出比特同时受当前输入位和前一个输入位的影响。第三位输入时，第一、二位分别右移一位，同时输出二个由这三位移位寄存器存储内容所共同决定的比特。当第四位输入时，第一位移出移位寄存器而消失。

移位过程可能产生的各种序列可以用图 9-8 所示的树状图来表示。树状图从节点 a 开始画，此时移位寄存器状态（即存储内容）为 00。当第一个输入位  $m_1=0$  时，输出位  $x_{1,1}x_{2,1}=00$ ；若  $m_1=1$  时，输出位  $x_{1,1}x_{2,1}=11$ ；因此从 a 点出发有两条分支（树叉）可以选择，也就是  $m_1=0$  时取上面一条分支， $m_1=1$  时取下面一条分支。当输入第二位时，移位寄存器右移一位后，在上分支情况下，移位寄存器的状态仍为 00，下分支的状态则为 01，把 01 状态记为 b。当新的一位输入时，随着移位寄存器状态和输入位的不同，树状图继续分叉成 4 条分支，2 条向上，2 条向

下。上分支对应于输入 0 状态，下分支对应于输入 1 状态。如此继续下去，即可得到图 9-8 所示的二叉树图形。树状图中，每条树叉上所标注的码元位输出状态，每个节点上标注的 a、b、c、d 表示移位寄存器的状态，也就是以前输入信息，a 状态表示  $m_{j-2}m_{j-1}=00$ ，b 状态表示  $m_{j-2}m_{j-1}=01$ ，c 状态表示  $m_{j-2}m_{j-1}=10$ ，d 状态表示  $m_{j-2}m_{j-1}=11$ 。显然，对于第  $j$  个输入位，就有  $2^j$  条分支，但是在  $j=N \geq 3$  时，树状图的节点自上而下开始重复出这 4 种状态。

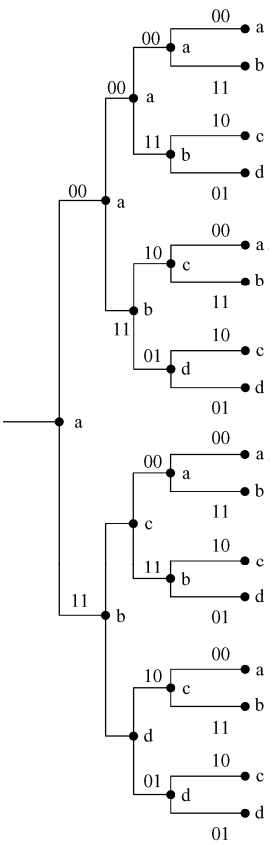


图 9-8 (2, 1, 3) 卷积码树状图

2. 网格图

从卷积码的树状图中观察中，可以看到树状图的节点自上而下会出现重复特性，为此，我们可以得到一种更为紧凑的图形表示方法，即网格图法，具体情况见图 9-9。

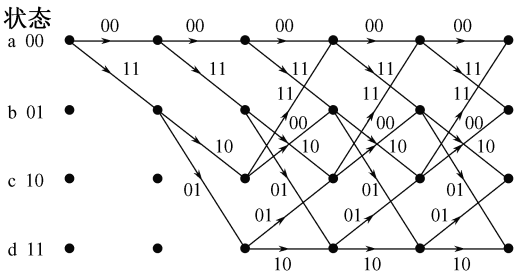


图 9-9 (2, 1, 3) 卷积码网格图



在网格图中，把码树中具有相同状态的节点合并在一起，码树中的上分支（对应输入 0），用实线表示，下分支（对应输入 1），用虚线表示。网格图中分支上标注的码元为对应的输出，自上而下 4 行节点分别表示 a、b、c、d 四种状态。一般情况下应有  $2^{N-1}$  种状态，从第  $N$  节开始，网格图图形开始重复而完全相同。

### 3. 状态图

观测图 9-8 可以看到，对于每一个节点当前状态 a、b、c、d，根据不同的输入将进入不同的状态，基于这一原理，我们可以构造出当前状态与下一状态之间的状态转换图，也可以称之为卷积码的状态图。在图 9-10 中实线表示信息位为 0 的路径，虚线表示信息位为 1 的路径，并在路径上写出相应的输出码元。当然，如果将状态图在时间上展开，便可以得到前面讲到的网格图。

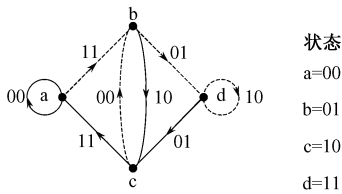


图 9-10 (2, 1, 3) 卷积码状态转换图

假如利用图 9-10 所示卷积码编码器，对输入序列 110111001000 进行编码，这时我们就可以利用上述 3 种方法当中任意一个，来分析编码器的输出序列和状态变化路径，这里以网格图为例进行分析。

若起始状态为 a，可以得到图 9-11 所示的结论。

通过上述分析以及对具体实例的研究，我们可以得到  $(n, k, N)$  卷积码的一些基本特性：

- (1) 对于每组  $k$  位的输入，利用卷积码编码后将得到  $n$  位的输出；
- (2) 树状图中每个节点可引出  $2^k$  条分支；
- (3) 网格图和状态图都有  $2^{k(N-1)}$  种可能的状态；每个状态可以引出  $2^k$  条分支，同时也有  $2^k$  条分支从其他状态或本状态引入；
- (4) 在任何情况下，只要卷积码编码器一确定，相应的树状图、网格图和状态图都将确定与输入的码序列无关。

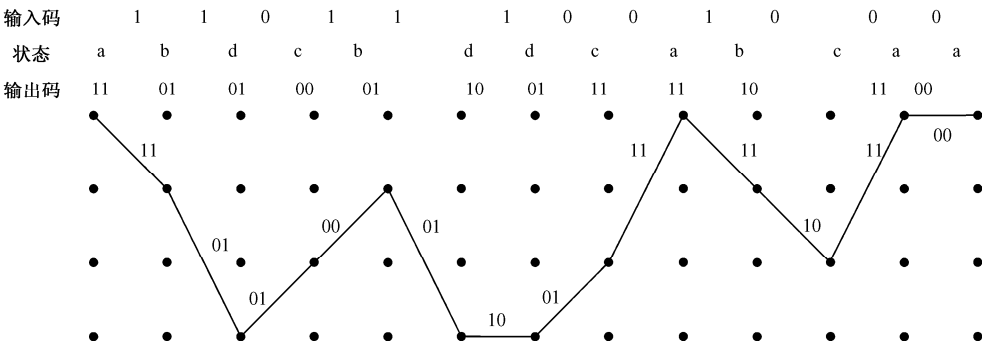


图 9-11 (2, 1, 3) 卷积码的编码过程及路径（网格图）

### 4. 延时算子多项式

在延时算子多项式表示方法中，编码器的移位寄存与模 2 和的连接关系以及输入、输出序列都表示为延时算子  $D$  的多项式。例如，输入序列 1101110...可表示为

$$M(D) = 1 + D + D^3 + D^4 + D^5 + \dots \tag{9-30}$$

其中变量  $D$  的幂次等于相对于时间起点的单位延时数目,时间起点通常选在第 1 个输出码元。在一般情况下,输入序列可表示为:

$$M(D) = m_1 + m_2D + m_3D^2 + m_4D^3 + \dots \quad (9-31)$$

这里,  $m_1, m_2, m_3, \dots$  表示输入码元 (1 或 0)。

用  $D$  算子多项式表示移位寄存器各级与各模 2 和连接关系时,若某级寄存器与某模 2 和相连,则多项式中相应项的系数为 1,否则为 0 (表示无连接线)。以图 9-6 所示 (2, 1, 3) 卷积码为例,左、右两个模 2 和与寄存器各级的连接关系可表达为

$$\begin{aligned} G_1(D) &= 1 + D + D^2 \\ G_2(D) &= 1 + D^2 \end{aligned} \quad (9-32)$$

通常把上述式子称为卷积码的生成多项式,可以用多项式相乘计算出卷积码的输出。以输入序列 1101110... 为例,可以得到

$$\begin{aligned} x_1(D) &= G_1(D) \cdot M(D) = (1 + D + D^2) \cdot (1 + D + D^3 + D^4 + D^5 + \dots) = 1 + D^5 + D^7 + \dots \\ x_2(D) &= G_2(D) \cdot M(D) = (1 + D^2) \cdot (1 + D + D^3 + D^4 + D^5 + \dots) = 1 + D + D^2 + D^4 + \dots \end{aligned} \quad (9-33)$$

对应输出序列为

$$\begin{aligned} x_1 &= (x_{1,1}, x_{1,2}, x_{1,3}, \dots) = 10000101\dots \\ x_2 &= (x_{2,1}, x_{2,2}, x_{2,3}, \dots) = 11101011\dots \\ x &= (x_{1,1}, x_{2,1}, x_{1,2}, x_{2,2}, x_{1,3}, x_{2,3}, \dots) = 1101010001100111\dots \end{aligned}$$

上述结果与图 9-11 种图解法所的结果相同。另外,为了方便地表示生成多项式,也可以采用类似循环码的表示方法,用二进制或八进制序列来表示生成多项式,例如式 (9-32) 就可以表示成为

$$G_1(D) \Rightarrow g_1 = (111)_2 = (7)_8 \text{ 和 } G_2(D) \Rightarrow g_2 = (101)_2 = (5)_8$$

显然,在编码效率及约束条件相同时,采用不同的生成多项式所得到的卷积码是不同的,如何选择生成多项式,以获得最佳的纠错能力,是一个非常重要的问题。

## 5. 二进制卷积码的距离特性

我们知道,在分组码中汉明码距与纠错能力有密切关系,生成一种分组码时应使码组之间的距离尽可能地大。常以最大的最小码距作为纠错能力的度量。

卷积码中也同样存在码距的概念,通常使用的码距有两种:最小码距  $d_{\min}$  和自由码距  $d_{\text{free}}$ 。

卷积码中长度为  $nN$  (假设约束长度为 7) 的编码后序列之间的最小汉明距离被称为最小码距  $d_{\min}$ ,任意长编码后序列之间的最小汉明距离被称为自由码距  $d_{\text{free}}$ 。由于卷积码并不划分码组,因而以自由码距作为纠错能力的度量更为合适。

在求  $d_{\min}$  或  $d_{\text{free}}$  时,我们并不需要列出所有可能出现的编码后序列。由于卷积码的线性性质,所有码序列之间的最小汉明距应等于非零码序列的最小汉明重量,即非零码序列中 1 码的个数。由此可见,当要求解  $d_{\min}$  或  $d_{\text{free}}$  时,只要考虑码树中下半部的码序列就可以了,因为这些码序列全部是以非零分支开始的。例如要求图 9-6 所示卷积码的最小距,由于约束长度  $N=3$ ,只需求树状图中下半部所有的 3 条分分支径的汉明重量就够了。共有 4 条包含这 3 条分支的路径: abca, abcb, abdc 和 abdd, 它们的汉明重量分别为 5, 3, 4, 4, 因而  $d_{\min}=3$ , 它所对应的编码后序列为 111000。

在求  $d_{\text{free}}$  时，必须对任意长度的路径加以考察。但由于  $d_{\text{free}}$  是有限的，而且由于路径的汉明重量在与全 0 路径汇聚以后已停止增长，于是与  $d_{\text{free}}$  对应的路径必然最后汇聚到全 0 路径上。因此  $d_{\text{free}}$  可以从由全 0 序列出发再回到全 0 序列的所有路径中求得，这可以用网格图或状态图来求。以图 9-11 所示的网格图为例，容易看出路径 abca 符合自由码距的定义， $d_{\text{free}}=5$ ，相应的输出码序列为 111011。

在上述例子中，与  $d_{\text{min}}$  和  $d_{\text{free}}$  对应的路径都只有一条，在另外一些卷积码中可以是多条。此外，随着路径长度的增加，码序列间的距离是非降的，因此一般来说  $d_{\text{min}} \leq d_{\text{free}}$ ，但对不少卷积码有  $d_{\text{min}}=d_{\text{free}}$ 。

应当指出，卷积码的好码（或称最佳码）的定义并不仅仅限于最大的自由码距，也可以定义为最小的码距，或最佳的距离分布特性，这取决于所用的译码器类型和应用场合。

### 9.3.2 卷积码的译码

卷积码的译码可分成代数译码和概率译码两类。

卷积码代数译码沿用了分组码代数译码的思路和概念，先从混合了发送码和噪声的接收序列  $B(D)=A(D)+E(D)$  中求出伴随式  $S(D)$ ，再根据  $S(D)$  确定错误图案  $E(D)$ ，在接收序列  $B(D)$  中减掉差错图案  $E(D)$ ，即可求出译码  $A(D)$ ；最后可获得发送信息  $M(D)$ 。其优点是译码电路简单，延时小，适合于高速译码。不足之处在于适合代数译码的卷积码的编码增益一般都不大，所以应用较少。

卷积码概率译码，在空间通信、无线通信中应用较多。概率译码依据于编码思路，对于编码器编出的任何码字序列，在网格图上一定可以找到一条连续的路径与之对应，这种连续性正是卷积码码字前后相关的体现。在接收译码端，一旦传输、或存储过程中出现差错，输入到译码器的接收码字流在网格图上就找不出一条对应的连续路径，而只有若干不确定、断续的路径供译码参考。从译码器译出的码字序列必须与编码器一样，也对应一条连续路径，否则肯定是译码出了错误。

概率译码是以断续的接收码流为基础，逐个计算它与其他所有可能出现的、连续的网格图路径的距离，选出其中可能性（概率）最大的一条作为译码估值输出。概率最大在大多数场合可解释为距离虽小，这种最小距离译码体现的正是最大似然的准则。

卷积码的概率译码就是搜遍网格图，找出最可能的序列。它既可以执行软判决或硬判决，因此，搜寻网格图时所用的量度可以是汉明距离，也可以是欧氏距离。在这里我们主要针对二进制对称信道进行讨论，重点对最大似然译码——维特比译码进行分析。

#### 1. 卷积码的最大似然译码

利用最大似然序列估计器就是使似然函数最大的一个参数作为估计值，鉴于似然函数的性质，我们通常选择似然函数的对数最大，而不是似然函数本身最大，现在我们就根据卷积码编译码系统的特点构造似然函数。

利用网格图表述卷积码的特点，设  $c_{jm}$  表示卷积编码后发送的数据信息， $j$  表示在状态转移过程中第  $j$  个分支（或称为状态）， $m$  表示该分支利用二进制表示时的第  $m$  位数，以图 9-6 表示的（2，1，3）卷积码为例，假如从初始状态  $a$  经过 3 次状态转移（3 个分支）又回到状态  $a$ ，观察网格图中的两条路径，与它们对应的信息序列分别是 000 和 100，对应的发送序列

分别是 00 00 00 和 11 10 11, 这时  $j=1, 2, 3, m=1, 2$ 。同样用  $\{r_{jm} \ j=1, 2, 3; m=1, 2\}$  表示卷积译码器的输入。穿过网格图可以有多条路径, 其中第  $i$  条路径的第  $j$  分支的量度通常表示为  $\{c_{jm}^{(i)}, m=1, 2\}$  而接收序列是  $\{r_{jm}, m=1, 2\}$ , 它们的联合条件概率的对数为

$$\mu_j^{(i)} = \lg P(Y_j / C_j^{(i)}), \quad j=1, 2, 3 \cdots \quad (9-34)$$

这样就可以将穿过网格图由  $B$  个分支组成的第  $i$  条路径的量度定义为

$$PM^{(i)} = \sum_{j=1}^B \mu_j^{(i)} \quad (9-35)$$

这就是似然函数对数的累加值, 现在我们就来确定似然函数的对数  $\mu_j^{(i)}$ 。如用软判决译码, 且序列用二进制相干 PSK 传输, 则译码器的输入为

$$r_{jm} = \sqrt{E_c} (2c_{jm} - 1) + n_{jm} \quad (9-36)$$

式中,  $n_{jm}$  表示加性高斯白噪声,  $E_c$  表示发送每个二进制码元所用的信号能量。则卷积译码器输入的联合条件概率密度函数为

$$p(r_{jm} | c_{jm}^{(i)}) = \frac{1}{\sqrt{2\pi}\sigma} \exp \left( -\frac{(r_{jm} - \sqrt{E_c} (2c_{jm}^{(i)} - 1))^2}{2\sigma^2} \right) \quad (9-37)$$

式中,  $\sigma^2 = N_0/2$  是加性高斯噪声的方差。如果忽略各分支量度相同拥有的项, 则似然函数的对数可以表示为

$$\mu_j^{(i)} = \sum_{m=1}^n r_{jm} (2c_{jm}^{(i)} - 1), \quad j=1, 2, 3 \cdots \quad (9-38)$$

从式 (9-38) 可以证明, 要使似然函数的对数  $\mu_j^{(i)}$  取最大值, 则  $r_{jm} = kc_{jm}^{(i)}$ , 其中  $k$  表示一个常数, 从物理上讲这个条件就是在  $r_{jm}$  和  $c_{jm}^{(i)}$  之间欧氏距离最小的情况下,  $\mu_j^{(i)}$  才能取最大值。

如用硬判决译码, 则卷积译码器的输出不是 0 就是 1。若各种概率满足关系:  $P(0/0)=P(1/1)$ ,  $P(1/0)=P(0/1)$ , 则该信道为二进制对称信道, 如果对译码器输出序列进行考察, 以判定长度为  $L$  的  $2^L$  个可能发送的序列中究竟是哪一个进入了编码器, 也就是确定输入信息序列  $M$ 。当某一特定的信息  $M$  进入编码器时, 发送序列为  $X(M)$ , 接收到的序列  $Y$ , 而条件概率  $P(Y/X(M))$  就被称为似然函数, 此似然函数对数为

$$\mu_j^{(i)} = \lg P(Y_j / X_j^{(i)}), \quad j=1, 2, 3 \cdots \quad (9-39)$$

假设发送序列  $X$  的长度为  $L$  个符号, 并在传输过程中发生了  $e$  个符号错误, 即  $X$  与  $Y$  有  $e$  个位置上符号不同, 它们的汉明距为  $e$ , 同时  $P(1/0)=P(0/1)=p$ , 这样对数似然函数为

$$\lg P(Y/X) = \lg(p^e (1-p)^{L-e}) = L \lg(1-p) - e \lg((1-p)/p) = -A - B \cdot e \quad (9-40)$$

对于  $p < 0.5$ , 则  $A$  和  $B$  均为正常数。因此, 汉明距  $e$  最小就相当于对数似然函数最大。这说明求最大似然函数的对数, 就相当于求  $X$  和  $Y$  之间的汉明距最小。

由此可见, 最大似然译码的任务就是在树状图或网格图中选择一条路径, 使相应的译码结果和输入码之间欧氏距离或汉明距离最小。

对于长度为  $L$  的二进制序列的最佳译码, 需要对可能发送的  $2^L$  个不同的序列的  $2^L$  条路径似然函数累加值 (即路径量度) 进行比较, 选取其中最大 (即最小量度) 的一条。显然,

译码过程的计算量随  $L$  增加而指数增长, 这在实际中难以实现, 因此只能采用次最佳的译码方法。

用网格图描述时, 由于路径的会聚消除了树状图中的多余度, 译码过程中只需考虑整个路径集合中那些能使似然函数最大的路径。如果在某一节点上发现某条路径已不可能获得最大对数似然函数, 那么就放弃这条路径。然后在剩下的“幸存”路径中重新选择译码路径, 这样一直进行到最后第  $L$  级。由于这种方法较早地丢弃了那些不可能的路径, 从而减轻了译码的工作量, 维特比译码正是基于这种想法。

## 2. 维特比译码

卷积码的网格图中共有  $2^{k(N-1)}$  种状态, 每个节点 (即每个状态) 有  $2^k$  条分支引入也有  $2^k$  条分支引出。为简便起见, 我们讨论  $k=1$  的情形, 从全 0 状态起始点开始讨论。由网格图的前  $N-1$  条连续分支构成的路径互不相交, 即最初的  $2^{(N-1)}$  条路径各不相同, 当接收到第  $N$  条分支时, 每条路径都有 2 条分支延伸到第  $N$  级上, 而第  $N$  级上的每两条分支又都会聚在一个节点上。

在维特比译码算法中, 把会聚在每个节点上的两条路径的对数似然函数累加值进行比较, 然后把具有较大对数似然函数累加值的路径保存下来, 而丢弃另一条路径, 经挑选后第  $N$  级只留下  $2^{(N-1)}$  条幸存路径, 选出的路径连同它们的对数似然函数累加值一起被存储起来。由于每个节点引出两条分支, 因此以后各级中路径的延伸都增大一倍, 但比较它们的似然函数累加值后, 丢弃一半, 结果留存下来的路径总数保持常数。由此可见, 上述译码过程中的基本操作是“加一比一选”, 即每级求出对数似然函数累加值, 然后两两比较并作出选择。有时会出现两条路径的对数似然函数累加值相等的情形, 在这种情况下可以任意选择其中一条作为“幸存”路径。

既然在每一级中都有  $2^{(N-1)}$  条幸存路径, 那么当序列发送完毕后, 如何判断其最后结果呢? 这就要在网格图的终结处加上  $N-1$  个已知信息 (即  $N-1$  条已知分支) 作为结束信息。在结束信息到来时, 由于每一状态中只有与已知发送信息相符的那条支路被延伸, 因而在每级比较后, 幸存路径减少一半。因此, 在接收到  $N-1$  个已知信息后, 在整个网格图中就只有唯一的一条幸存路径保留下来, 这就是译码所得的路径。也就是说, 在已知接收到的序列情况下, 这条译码路径和发送序列是最相似的。

由上述可见, 维特比译码器的运行是前向的、无反馈的。由于每级中每个状态上要进行“加一比一选”运算, 因此译码器的复杂性与状态数成正比, 因而也是随约束长度  $N$  的增加而指数增长的。因此目前只限于应用在较短约束长度 ( $N \leq 10$ ) 的卷积码中。

需要指出的是, 上述作为结束信息为确知信息。因此, 只要差错模式不超出卷积码的纠错能力, 从一个节点开始分叉产生的各条幸存路径经过一段间隔后总能正确地又合并成一条路径。但需经过多长间隔, 在何处合并, 都是不肯定的, 这与差错模式有关。显然, 在实际实现时, 不可能建立这种随机的译码深度, 而只能建立一个固定的译码深度。显然, 译码深度  $M$  和状态数  $2^{(N-1)}$  决定了需要存储的内容, 因为在路径合并成一条之前长度为  $M$  的  $2^{(N-1)}$  条路径必须全部保存起来。只有当具有最大似然函数累加值 (即最小量度) 的路径判别后才能将存储器内容刷新。因此存储容量至少为  $M \times 2^{(N-1)}$  个量度和分支。译码深度  $M$  通常是用计算机模拟来确定的, 在性能和设备量之间取一个折中。

译码深度实际上是译码器的约束长度，它也是译码器所产生的译码延时。表 9-5 中给出不同译码约束长度和不同编码约束长度时的硬判决维特比译码的  $E_b/n_0$  性能，这些数据是在采用 PSK 调制；干扰为加性白色高斯噪声，误信率  $P_b=10^{-5}$  的条件下由计算机模拟得到的。表 9-6 给出 3 比特软判决维特比译码时的  $E_b/n_0$  性能。

表 9-5 硬判决维特比译码的  $E_b/n_0$ (dB)( $P_b=10^{-5}$ ,  $n=2$ ,  $k=1$ )

| <div> <div>编码约束长度</div> <div><math>E_b/n_0</math></div> <div>译码约束长度</div> </div> |  | 5    | 6   | 7   |
|--|--|------|-----|-----|
| 3  |  | 7.4  | 7.2 | 6.7 |
| 4  |  | 7.25 | 6.8 | 6.5 |
| 5  |  | 7.2  | 6.7 | 6.4 |
| 10   |  | ---  | --- | 6.3 |

表 9-6 3 比特软判决维特比译码的  $E_b/n_0$ (dB)( $P_b=10^{-5}$ ,  $n=2$ ,  $k=1$ )

| <div> <div>编码约束长度</div> <div><math>E_b/n_0</math></div> <div>译码约束长度</div> </div> |  | 5   | 6   | 7   |
|--|--|-----|-----|-----|
| 3  |  | 5.7 | 5.2 | 5.1 |
| 4  |  | 5.3 | 5.0 | 4.7 |
| 5  |  | 5.2 | 4.9 | 4.5 |
| 27   |  | --- | --- | 4.4 |

由表 9-5、9-6 可知，当译码约束长度与编码约束长度相接近时，无论是硬判决或软判决维特比译码，它们所带来的  $E_b/n_0$  改善已接近饱和。因此译码约束长度不必取得很长，这对于硬件实现是有利的。在实际应用当中，通常取译码约束长度为编码约束长度的 3~5 倍译码足够了。

为了具体地阐明维特比译码的过程，我们仍以图 9-6 所示的 (2, 1, 3) 卷积码为例进行分析。假设卷积编码器输出为  $X=000000000000$ ，而维特比译码器的输入  $Y=001001000000$ ，因此上述接收序列是误码序列，为了使全部信息通过译码器，在卷积编码器输入加 3 个零，这时对应的卷积编码器输出为  $X=000000000000000000$ ，而译码器的输入  $Y=001001000000000000$ 。图 9-12 表示随着接收序列的串行输入维特比译码器中各条路径的取舍情况。对于硬判决译码，对数似然函数可以用计算该路径与接收序列之间的汉明距离来代替。图 9-12 中节点处所引入的分支上圆圈内的数字，它代表从起始点到某节点的该路径与接收序列之间的汉明距离，即路径量度。图 9-12 中 (1) 到 (10) 分别画出了每输入 1 比特信息时网格图中路径的变化。

图 9-12 (1) 表示从起始点出发可能出现的两条分支。上分支相应于输入信息为“0”，卷积编码后输出信息为“00”；下支路相应于输入信息为“1”，卷积编码后输出信息为“11”；两条支路的输出比特与接收序列中相应比特的汉明距离被称为分支重量，分别为 0 和 2。

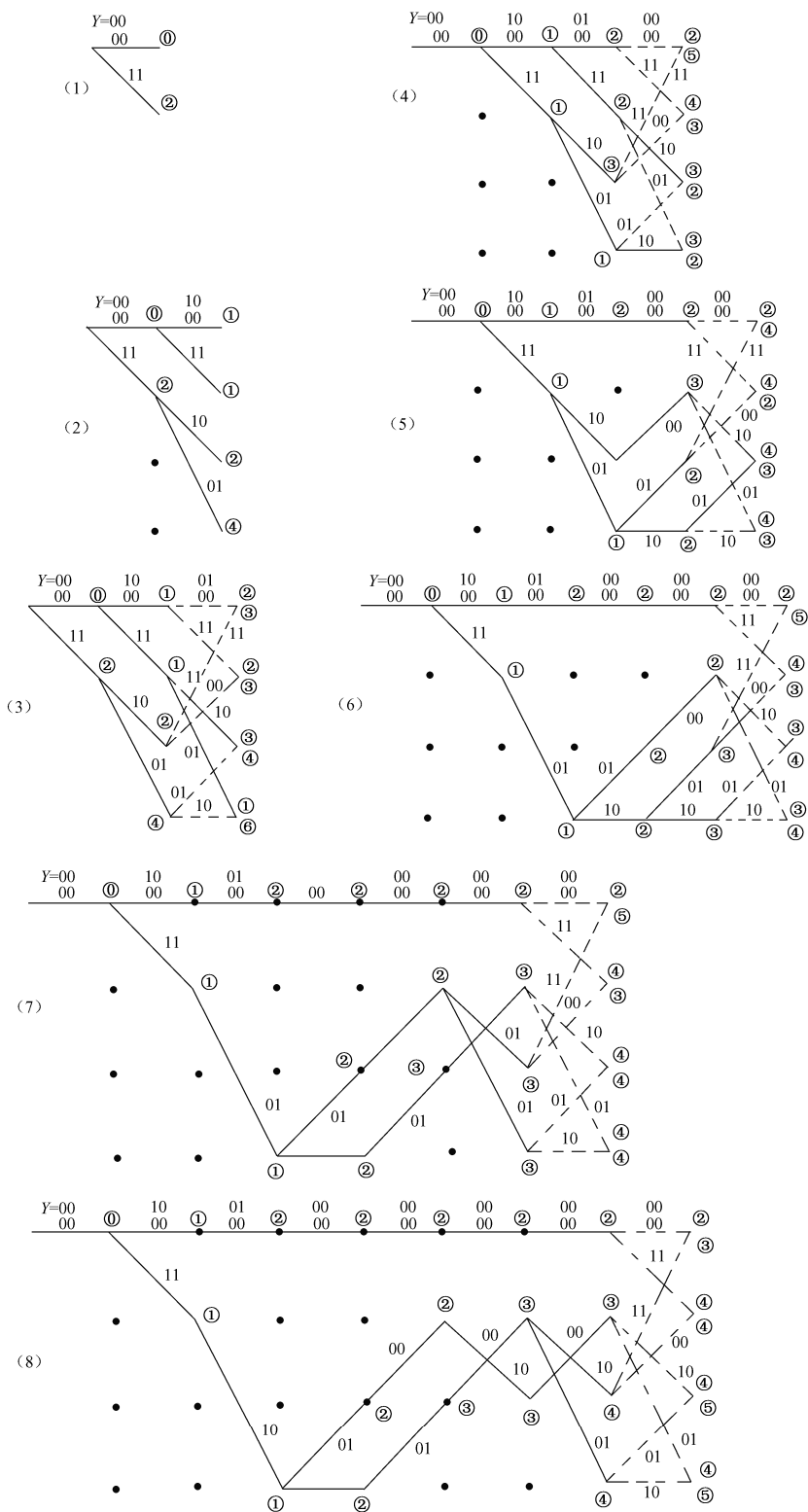


图 9-12 维特比译码过程的网格图表示

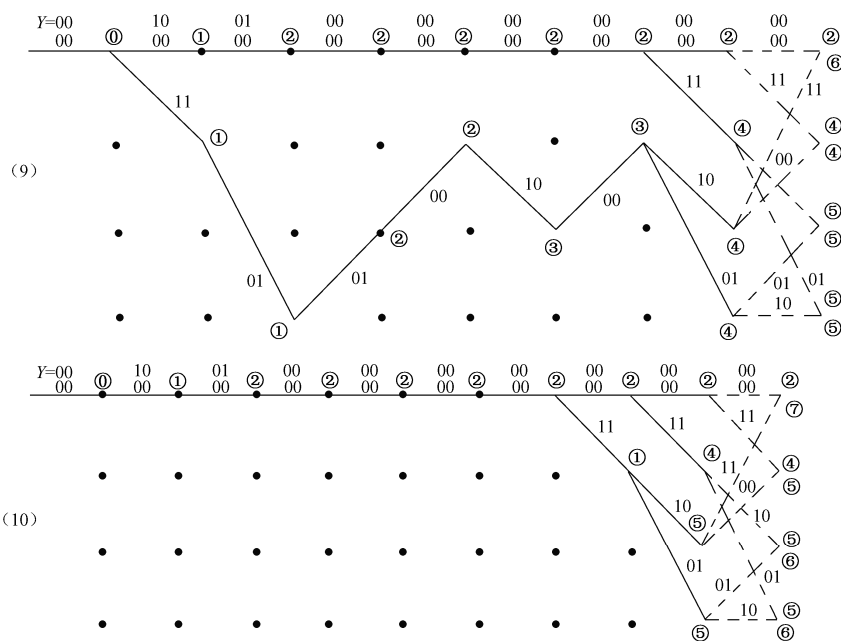


图 9-12 维特比译码过程的网格图表示 (续)

图 9-12 (2) 中, 第二级出现了 4 条互不会聚的分支, 这 4 条分支连同与它们相连的前面两分支, 共同构成了四条路径, 它们的路径重量为 1、1、2、4。

图 9-12 (3) 中, 第三级网格共出现 8 条支路, 这时对进入每个节点的二条路径的量度 (也就是前面所讲的路径重量) 进行比较, 将量度小的一条路径保留下来, 而丢弃量度大的路径。因而, 幸存路径只有 4 条, 它们的路径量度分别为 2、2、3、1。

进入到第四级网格时这 4 条幸存路径又延伸为 8 条, 经计算路径量度并比较后, 又丢弃其中的 4 条。遇到量度相同的情况, 可任意丢弃一条路径。如此继续下去, 在图 9-12 (10) 中可以看到, 所有幸存路径已全部合并成一条全 0 路径, 即纠正了接收序列中的误码。表 9-7 给出与图 9-12 相对应的维特比译码器中存储器内的信息, 这就是译码器可能的输出序列。

表 9-7 维特比译码器存储器内容

| 译码步骤 | 路径量度 | 存储器内容 | 译码步骤 | 路径量度 | 存储器内容    |
|------|------|-------|------|------|----------|
| 1    | 0    | 0     | 6    | 2    | 000000   |
|      | 2    | 1     |      | 3    | 011010   |
|      |      |       |      | 3    | 011011   |
|      |      |       |      | 3    | 011101   |
| 2    | 1    | 00    | 7    | 2    | 0000000  |
|      | 1    | 01    |      | 3    | 0110101  |
|      | 2    | 10    |      | 4    | 0111010  |
|      | 4    | 11    |      | 4    | 0111011  |
| 3    | 2    | 000   | 8    | 2    | 00000000 |
|      | 2    | 001   |      | 4    | 00000001 |
|      | 3    | 010   |      | 4    | 01101010 |
|      | 1    | 011   |      | 4    | 01101011 |



续表

| 译码步骤 | 路径量度 | 寄存器内容 | 译码步骤 | 路径量度 | 寄存器内容     |
|------|------|-------|------|------|-----------|
| 4    | 2    | 0000  | 9    | 2    | 000000000 |
|      | 3    | 0101  |      | 4    | 000000001 |
|      | 2    | 0110  |      | 5    | 000000010 |
|      | 2    | 0111  |      | 5    | 000000011 |
| 5    | 2    | 00000 |      |      |           |
|      | 2    | 01101 |      |      |           |
|      | 3    | 01110 |      |      |           |
|      | 3    | 01111 |      |      |           |

事实上，表 9-7 中在第 3 步即可断定第一条支路应为  $Y=00$  支路，因为所有幸存路径都是从同一点出发的。同样，在第 6 步可断定第二条支路也是  $Y=00$  支路，因为这条路径的量度最小。由此可见，不必等到最后得到唯一一条路径才作出最后的判断。译码器可以以少量时延连续不断地工作，作出基本上是正确的判断。

显然，维特比译码并不能纠正所有可能发生的错误，当错误模式超出卷积码的纠错能力时，译码后的输出序列就会带有错误。上例中，在长度为 2 条支路的接收序列中发生了 2 个错误，而我们知道这个卷积码的自由距为 5（路径长度为 3 时的最小汉明距），因而是可以纠正的。如果在路径长度  $\leq 3$  时，出现了 3 个错误，则这种错误模式已超出该码的纠错能力，而成为不能纠正的了。

3. 维特比译码器的实现

图 9-13 表示维特比译码器方框图，它包括以下各部分：

（1）输入与同步单元。它为译码器提供正确的支路同步，每次正确地输出属于一条支路的  $n$  个比特。

显然，当支路定时失步时，译码过程中将会出现大量的差错，只要能检测出这种状态，即能有效地调整支路同步。一种方法是监视路径量度的增长率，另一种方法是检查网格图的路径合并性质。当译码器出现失步时，网格图中幸存路径合并的速率比同步时慢得多。

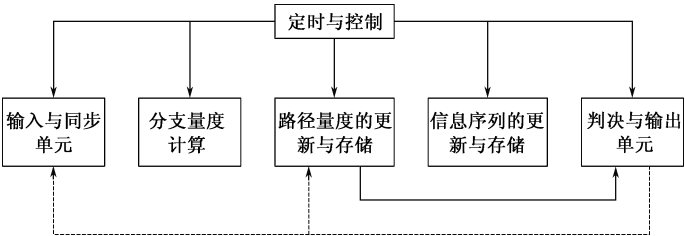


图 9-13 维特比译码器原理方框图

（2）支路量度计算。每当接收到一条新支路的一组  $n$  个量度值（硬判决时为  $n$  比特），支路量度计算单元就对网格图中每一条不同的支路确定一新的量度值。对于编码效率为  $k/n$  的卷积码，每次将有  $2^n$  个不同的量度值。在软判决维特比译码时，支路量度值不但随支路不同而异，而且还与接收信号的量化值有关。

（3）路径量度的存储与更新。在此单元中，支路量度与以前所存储的路径量度相加，然

后对会聚到同一节点处的支路进行路径量度比较, 选择一条路径量度最小的路径保留下来。这单元中的关键电路是“加一比一选”电路。

(4) 信息序列的存储与更新。一种最佳也是最常用的方法是基于最大似然译码。对于编码效率为  $1/n$  卷积码而言, 每接收一组新的支路信息, 在  $2^{(N-1)}$  个状态的路径存储器中存入经“加一比一选”电路选出的一位假想信息比特, 同时将最先存入路径存储器的一位比特输出给判决单元。因此, 每接收到一条新支路,  $2^{(N-1)}$  个路径存储器就更新一次它所存储的假想信息序列。路径存储器长度即为译码深度。当采用并行的处理方法时, 这种方案所需的路径存储量及“加一比一选”电路硬件数量很大; 也可以采用其他简化的方法来实现, 但在工作速度或性能方面要受影响。

(5) 判决与输出单元。在编码效率为  $1/n$  卷积码最佳译码时, 应选择具有最小路径量度的假想信息序列中最早存入的一个比特做译码输出。在确定最小量度路径时, 至少要作  $2^N-1$  次比较。

如上所述可知, 维特比译码器的硬件量随着约束长度  $N$  增大而迅速增加, 但近年来, 随着超大规模集成电路的发展。现在已能制造出单片维特比译码器, 并以开始广泛应用于卫星通信当中。

## 9.4 网格编码调制

### 9.4.1 引言

信道的纠错编码需要冗余度, 编码增益是依靠降低信息传输率来获得的。在功率受限信道中, 功率利用率可以用频带利用率换取。在限带信道中, 则可通过加大调制信号集来为纠错编码提供所需的冗余度, 以避免信息传输速率因加纠错编码而降低。但若调制和编码仍以相互独立的方法设计, 则不能得到令人满意的结果。

例如, 有一个系统用不进行纠错编码的 4PSK 调制, 在差错概率  $10^{-6}$  处可以得到  $2 \text{ b} \cdot \text{s}^{-1} / \text{Hz}$  的频带利用率, 在这个差错概率条件下, 每比特所需的 SNR 是  $\gamma_b=10.5 \text{ dB}$ 。在不能扩大带宽的条件下, 如果想利用信道编码来减小  $\gamma_b$ , 这里假设选用编码效率  $R_c=2/3$  的码, 这样的编码必然结合 8PSK 那样的调制使用, 才能得到与不编码 4PSK 同样的数据吞吐量。然而, 信号相位数由 4 增加到 8 需要增加约 4 dB 信号功率才能保持同样的差错概率。所以, 要使编码有意义,  $R_c=2/3$  码的编码增益必须大到足以抵偿 4 dB 损失才行, 这对于编码效率为  $2/3$  信道编码来讲, 是一个相当大的编码增益, 实现起来较为困难。出现上述结果, 是由于在纠错编译码和调制解调部分的设计中存在下面两个问题。

一个问题是上述编码 8PSK 系统的解调器是将信号作独立硬判决, 硬判决结果再送给译码器译码。这种硬判决会导致接收端接收信息的不可恢复的丢失。解决这个问题的方法是在收端采用软判决译码, 这就意味着译码器将直接对调制信道的无限量化“软”输出抽样进行处理。设输出抽样为

$$r_n = a_n - \omega_n \tag{9-41}$$

这里  $a_n$  是输入调制器的离散信号,  $\omega_n$  代表加性高斯白噪声过程的抽样。此时最佳译码器的判决准则是从所有编码信号序列中选出一个序列  $\{\hat{a}_n\}$ , 使得序列  $\{\hat{a}_n\}$  和接收序列  $\{r_n\}$  之间具

有最小的平方欧氏距离，也即找出满足下式的序列 $\{\hat{a}_n\}$ ：

$$\sum |r_n - \hat{a}_n|^2 = \min \sum |r_n - a_n|^2 \quad (9-42)$$

当采用卷积码时，可用维特比译码算法来决定最靠近接收序列 $\{r_n\}$ 的码信号序列 $\{\hat{a}_n\}$ ，软判决维特比译码是一种最大似然译码。此时，所谓的“纠错”已不再是纠正某些解调“硬”判决结果，而是直接从软量化解调输出中确定最大似然编码序列，也就是确定与序列 $\{r_n\}$ 具有最小的平方欧氏距离的序列。

另一个问题是出现在判决量度上的差别。传统的纠错编码是以汉明距离作为量度进行设计的，但是对汉明距离而言为最佳的编码符号，在映射成非二进制调制信号时，并不能保证能获得好的欧氏距离结构。实际上，仅在 2PSK 和 4PSK 调制时，汉明距离才与欧氏距离等价，此时按汉明距离最佳而设计的码就欧氏距离而言也是最佳的。而在一般的多进制调制中，汉明距离和欧氏距离之间不存在单调的关系。

根据上面的讨论，可以看到最佳的编码调制系统应按编码序列的欧氏距离为调制设计的量度。这就要求必须将编码器和调制器当做一个统一的整体进行综合设计，使得编码器和调制器级联后产生的编码信号序列具有最大的欧氏自由距离。

从信号空间的角度看，这种最佳编码调制的设计实际上是一种对信号空间的最佳分割。这正是近几年来发展起来的网格编码调制（Trellis Coded Modulation, TCM）技术的基本原理。对于限带信道，网格编码调制是由昂格尔博克（Ungerboeck）于 1982 年最先提出的，它有两个最基本的特征：

- （1）星座图中所用的信号点数大于未编码同种调制所需的点数（通常扩大 1 倍），这些附加的信号点为纠错编码提供冗余度；
- （2）采用卷积码在相继的信号点之间引入某种依赖性，因而只有某些信号点序列才是允许出现的，这些允许的信号序列可以模型化为网格结构，因而称为网格编码调制。

### 9.4.2 集合分割原理

通常 TCM 最优码是按照编码信号的网格图确定的。当一个 TCM 最优码确定后，若要实现该 TCM 码相应的编码调制器，可有两种方法：一种方法是先确定从编码符号到调制信号的映射函数，再根据网格图设计出相应的纠错编码器；另一种方法是先确定编码器结构，再根据网格图确定编码符号到调制信号的映射函数。

在设计和选择 TCM 最优码方案的网格图结构时，通常赋予信号的网格转移以这样一种特性：即 TCM 方案通过一种特殊的信号映射可变成卷积码的形式。这种映射的原理是将调制信号集分割成子集，使得子集内的信号间具有更大的空间距离（欧氏距离），这也就是下面要讲的集分割原理。用这种方法构造的 TCM 方案可用图 9-14 所示的通用 TCM 编码调制器结构来解释。

根据这个结构，TCM 信号的产生过程如下：有  $k$  比特待传输信息 $\{m_i^k\}$ ，其中的 $\tilde{k}$ 比特（ $\tilde{k} < k$ ）通过一编码效率为 $\tilde{k}/(\tilde{k}+1)$ 的二进制卷积编码器扩展成 $(\tilde{k}+1)$ 编码比特，这 $(\tilde{k}+1)$ 个编码比特用来选择 $2^{\tilde{k}+1}$ 进制调制信号集的 $2^{\tilde{k}+1}$ 个子集中的一个，剩下的 $(k-\tilde{k})$ 个未编码比特用来选择传送该子集的 $2^{k-\tilde{k}}$ 个信号中的某一个。

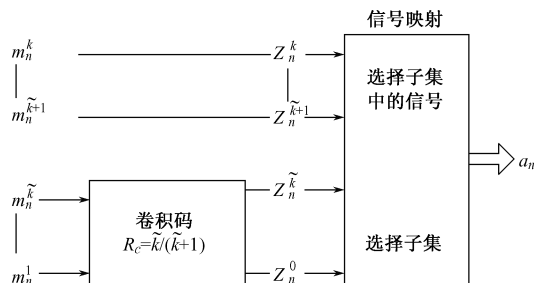


图 9-14 TCM 编码调制器结构

在图 9-14 中信号的映射是实现 TCM 调制的关键，而这个映射采用的是集分割原理，因此，集分割原理在 TCM 方案的构造中具有十分重要的意义。所谓集分割是将一信号集接连分割成较小的子集，并使分割后的子集内的最小空间距离得到最大的增加。每一次分割都是将一较大的信号集分割成较小的两个子集，这样可得到一个表示集分割的二叉树。每经过一级分割，子集数就加倍，而子集内最小距离亦增大。

设经过  $i$  级分割后子集内最小距离为  $\Delta_i$  ( $i = 0, 1, \dots$ )，则有  $\Delta_0 < \Delta_1 < \Delta_2 < \dots$  设计 TCM 时，将调制信号集作  $(\tilde{k} + 1)$  分割，直至  $\Delta_{\tilde{k}+1}$  大于所需的自由距离  $d_{\text{free}}$  为止。

在集分割树中，我们令始于第  $i$  级的同一结点的两个分支所对应的编码比特为  $Z_n^i = 0$  或  $1$ 。图 9-15 展示了 8PSK 集分割示意图。

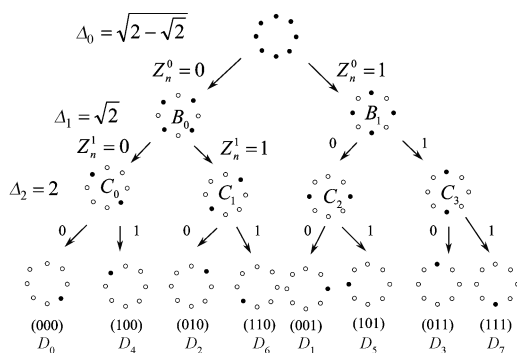


图 9-15 8PSK 集分割示意图

在共有  $(\tilde{k} + 1)$  级的集分割树中， $2^{\tilde{k}+1}$  个子集对应于不同的  $(\tilde{k} + 1)$  个编码比特： $Z_n^{\tilde{k}}, \dots, Z_n^0$ 。所以，此  $(\tilde{k} + 1)$  个编码比特  $Z_n = \{Z_n^{\tilde{k}}, \dots, Z_n^0\}$  将唯一地确定对应的子集，故可将  $Z_n$  看做该子集的标值。根据这种赋予子集的标值的方法，可以得出一个重要的性质：如果两个子集的标值的最末  $q$  位相同，但  $Z_n^q$  位不同，则两子集的信号是集分割树上第  $q$  级的同一子集的元素，因此两个子集的任两个信号之间的最小距离至少是  $\Delta_q$ ，这一性质称为集分割原理。

$(k - \tilde{k})$  个未编码比特用来从被选中的子集中选择其中一个信号。具体来说，这  $(k - \tilde{k})$  比特与子集中信号如何映射，在 TCM 码的设计中通常并不重要，因为它不影响 TCM 码的自由距离，故对码的性能影响不大。在网格图中，子集内的  $2^{k-\tilde{k}}$  个信号对应着  $2^{k-\tilde{k}}$  条并行转移支路。图 9-15 中画出了 8PSK 信号集的集分割二叉树，它的星座图示于图 9-16 中，该图中圆

的半径为 1，信号点旁的数字为信号点的标号。 $\Delta_0, \Delta_1, \Delta_2, \dots$  分别表示不同信号点之间的距离。

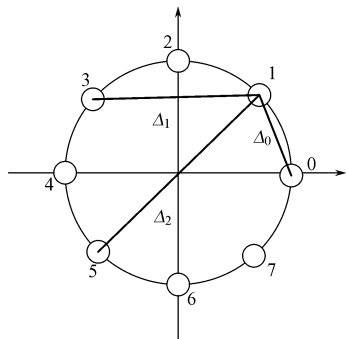


图 9-16 8PSK 星座图

### 9.4.3 TCM 码网格图的构造

有了上述信号点的子集划分后，当卷积编码器给定时，剩下的问题是如何进行映射，使调制信号之间欧氏距离最大，下面以 8PSK 为例,讨论 TCM 码的网格图的构造规则。

假设选用编码效率  $R_c=2/3$  的码，这样的编码必然结合 8PSK 那样的调制使用，才能得到与不编码 4PSK 同样的数据吞吐量。因此，在这里考虑 8PSK 信号星座与 TCM 码结合使用，把不编码的 4PSK 作为参照以便衡量编码增益。

不编码 4PSK 所用的信号点集为图 9-15 所示的  $B_0$  或  $B_1$ ，信号点间的最小距离  $\Delta_1 = \sqrt{2}$ 。这里假设对于编码的 8PSK 调制，采用如图 9-17 (a) 所示的四状态网格图进行编码。注意，网格图上的每一分支对应 4 个子集  $C_0$ 、 $C_1$ 、 $C_2$  或  $C_3$  之一，对于八点星座，每子集  $C_0$ 、 $C_1$ 、 $C_2$  和  $C_3$  包含 2 个信号点。因此，状态转移  $C_0$  包含的 2 点对应于比特 (000, 100)，用八进制表示为 (0, 4)。同样， $C_2$  包含的 2 点对应于 (010, 110) 或八进制 (2, 6)， $C_1$  包含的 2 点对应于 (001, 101) 或八进制 (1, 5)， $C_3$  包含的 2 点对应于 (011, 111) 或八进制 (3, 7)。这样，四状态网格图上的每个转移包含 2 条并行路径，如图 9-17 (b) 所示，可以看到，任何由同一状态发出，经过一次以上转移又在同一状态汇合的两条路径，它们之间欧氏距离的平方之和至少是  $\Delta_0^2 + \Delta_1^2 + \Delta_1^2 = 4.585$ ，例如，(0、0、0) 和 (2、1、2) 就是这样的两条信号路径。另一方面，并行转移欧氏距离平方之和是 4，因此，在四状态网格图中由任何状态发出又汇合于同一状态的两条路径之间的最小欧氏距离是 2，TCM 码中的这种最小距离叫做自由欧氏距离，用  $D_{fed}$  表示。将不编码的 4PSK 调制的欧氏距离  $\Delta_1 = \sqrt{2}$  与 8PSK 的四状态 TCM 码相比，8PSK 的四状态 TCM 码取得了 3 dB 的编码增益。

应该指出，对于图 9-17 (a) 所示的四状态 TCM 码，从它能提供最大的自由欧氏距离这个义上说是最优的。当然，还可以构成许多其他的四状态 TCM 码，但是其他可能的四状态 TCM 码都不可能给出更大的  $D_{fed}$ 。

最优的八点星座四状态 TCM 码的结构应符合下列带有普遍性的规则：

(1) 并行转移对应到具有最大欧氏距离间隔的信号点，比如 8PSK 的 4 个子集  $C_0$ 、 $C_1$ 、 $C_2$  和  $C_3$  的  $\Delta_2 = 2$ 。

(2) 对于许可构成 TCM 码的任意状态, 如子集  $(C_0C_2)$  和  $(C_1C_3)$  等, 它们之间具有的最大距离是  $\Delta_1=\sqrt{2}$ ;

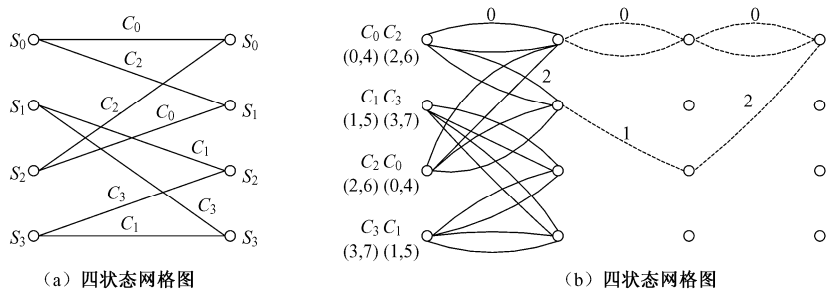


图 9-17 编码的 8PSK 调制

(3) 信号点应以同样的频度(次数)出现。

值得注意的是, 规则(1)和(2)保证从任何状态发出并汇合于该状态的单条或多条路径的欧氏距离一定大于不编码 4PSK 的欧氏距离。规则(3)保证 TCM 码一定有一个规则的结构。

应该指出, 编码比特具体如何映射到如图 9-15 所示的信号点集并不重要, 在该图上, 8 个信号点是用等效的二进制形式表示的。只要保证子集间具有增大的最小距离这一主要特征, 就完全可以改变子集而采用另一种映射。

### 9.4.4 TCM 码的卷积码表示

TCM 的最优码可有多种等效的网格图形式。图 9-17 中所示的网格图通过一定的映射方式(即按集分割原则的映射), 可将 TCM 码表示成一个编码效率为  $\tilde{k}/(\tilde{k}+1)$  的二进制卷积编码器和符合集分割原则的映射函数的级联, 也即如图 9-14 所示的结构。

$\tilde{k}/(\tilde{k}+1)$  卷积编码器将  $\tilde{k}$  比特输入经编码产生  $(\tilde{k}+1)$  比特输出  $Z_n=\{Z_n^{\tilde{k}},\cdots,Z_n^0\}$ 。此线性卷积编码器可用监督方程定义, 监督方程给予长度为  $v+l$  的滑动时间窗内的码元以一定的约束关系, 即: 使其满足等式

$$\sum(h_v^iZ_{n-v}^i\oplus\cdots\oplus h_{v-l}^iZ_{n-v+l}^i\cdots\oplus h_0^iZ_n^i)=0 \tag{9-43}$$

这里  $v$  为记忆长度,  $h_l^i$  值  $(0<l<v, 0<i<k)$  是码的监督方程的系数, 故整个 TCM 码的监督矩阵可表示为

$$\begin{aligned} H^j(D)&=0 & \tilde{k}<j<k \\ H^j(D)&=h_v^jD^v+h_{v-1}^jD^{v-1}+\cdots+h_1^jD+h_0^j & 0<j<\tilde{k} \end{aligned} \tag{9-44}$$

图 9-18 画出了与图 9-17 对应记忆长度为 2 的卷积编码器, 其监督矩阵和生成矩阵可以表示为

$$v=2: H(D)=[0 \quad D \quad D^2+1], G(D)=\begin{bmatrix} 1 & 0 & 0 \\ 0 & D^2+1 & D \end{bmatrix} \tag{9-45}$$

因为  $H(D)$  中的最高次幂等于  $v$ , 也就是 2, 故这是一个最小编码器。通常 TCM 方案的最优码是通过手算或计算机搜索的方法获得的。对一些状态数较小的常见 TCM 码, 人们已找出了其最优码。

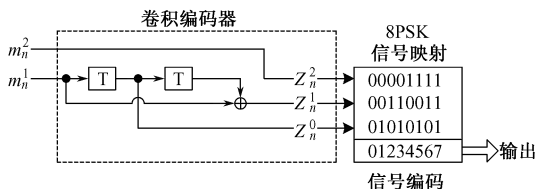


图 9-18 四状态网格最优码编码原理结构图

## 9.5 Turbo 码

Turbo 码首先由 C. Berrou、A. Giavieux 等人在 1993 年提出的一种新型信道编码方案，它巧妙地将卷积码和随机交织器结合在一起，实现了随机编码的思想，如果译码方式和参数选择得当，其性能可以接近 Shannon 极限。因此，这一超乎寻常的优异性能，立即引起信息与编码理论界的轰动。由于 Turbo 码的上述优异性能不是从理论研究角度给出的，仅是计算机仿真的结果，因此，Turbo 码的理论基础还不完善。随着研究的不断深入，其性能也被不断提高，并在强干扰环境下显示其广阔的应用前景。

目前，Turbo 码已经成为以大容量、高数据率和承载多媒体业务为目的的第三代移动通信的信道编码方案之一，其中，具有代表性的 WCDMA、CDMA 2000 和我国的 TD-SCDMA 这 3 个标准中的信道编码方案都使用了 Turbo 码。同时，由于 Turbo 码受到各国研发者的关注和重视，因此，发展非常迅速，已经开始用于深空通信、卫星通信及多媒体通信等领域。Turbo 码在高清清晰度数字电视传输系统的应用中，能使大量的数字信号准确无误地传输，真正做到高清晰度的图像质量。不仅如此，迭代译码的思想已经作为“Turbo 原理”而广泛地应用于均衡、调制、信道检测等领域。

### 9.5.1 编码结构和原理

Turbo 码最初以并行级联卷积码（Parallel Concatenated Convolutional Codes, PCCC）形式出现，后来为了克服误码率的错误平台，S. Benedetto 和 D. Divsalar 等人提出了串行级联卷积码（Serial Concatenated Convolutional Codes, SCCC），又称为串行级联 Turbo 码，将 PCCC 与 SCCC 相结合，S. Benedetto 设计了混合级联卷积码（Hybrid Concatenated Convolutional Codes, HCCC）。鉴于理论分析不尽完善，实现过程过于复杂，下面将仅将 PCCC 的编码和译码原理进行简单的分析和介绍。

PCCC 形式编码器是由 2 个或者多个二元带反馈的递归系统卷积编码器（Recursive Systematic Convolutional, RSC），通过一个随机交织器并行级联而成，编码后的校验位可以按照要求经过删余矩阵进行删减，从而产生不同码率的码字。一个典型 Turbo 码的编码器结构如图 9-19 所示。

信息序列  $u = \{u_1, u_2, \dots, u_N\}$  经过一个  $N$  位交织器，形成一个新序列  $u_1 = \{u'_1, u'_2, \dots, u'_N\}$ ，这 2 个序列的长度与内容都相同，只是比特位置经过了重新排列。 $u$  与  $u_1$ ，分别传送到 2 个分量码编码器。通常这 2 个分量编码器结构相同，生成序列  $X^{P1}$  和  $X^{P2}$ 。为了提高码率，序列  $X^{P1}$  和  $X^{P2}$  需要经过删余器，从 2 个校验序列中周期地删除一些校验位，形成校验位序列

$\mathbf{X}^p$ 。  $\mathbf{X}^p$  与未编码序列  $\mathbf{X}^s$  经过复用后，生成 Turbo 码序列  $\mathbf{X}$ 。假设 2 个分量编码器的码率都是 1/2，那么就可以采用删除矩阵  $\mathbf{P}$  删除来自 RSC1 的校验序列  $\mathbf{X}^{p1}$  的偶数位置比特和来自 RSC2 的校验序列  $\mathbf{X}^{p2}$  的奇数位置比特。这样就可以得到码率为 1/2 的 Turbo 码序列。其中删除矩阵可以表示为

$$\mathbf{P}=\begin{Bmatrix} 1 & 0 \\ 0 & 1 \end{Bmatrix}$$

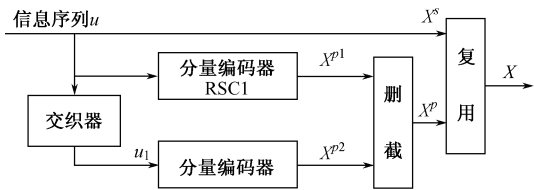


图 9-19 Turbo 码编码器结构图

### 9.5.2 译码器结构和原理

PCCC 译码器基本结构如图 9-20 所示，它由 2 个软输入软输出（Soft In Soft Out，SISO）译码器 DEC1 和 DEC2 串行级联组成，交织器与编码器中所使用的交织器相同。译码器 DEC1 对分量码 RSC1 进行最佳译码，产生关于信息序列  $\mathbf{u}$  中每一比特的似然信息，并将其中的“新信息”经过交织器送给 DEC2，译码器 DEC2 将此信息作为先验信息，对分量码 RSC2 进行最佳译码，产生关于交织后的信息序列中每一比特的似然比信息，然后将其中的“外信息”经过解交织器送给 DEC1，进行下一次译码。这样，经过多次迭代，DEC1 和 DEC2 的外部信息对于降低误比特率的作用逐渐减小，外部信息的值趋于稳定，似然比渐近一个稳定值，译码过程逼近于对整个码的最大似然译码，然后对此似然比进行硬判决，即可得到信息序列  $\mathbf{u}_k$ 。每一比特的最佳估计值序列  $\hat{\mathbf{u}}_k$ 。

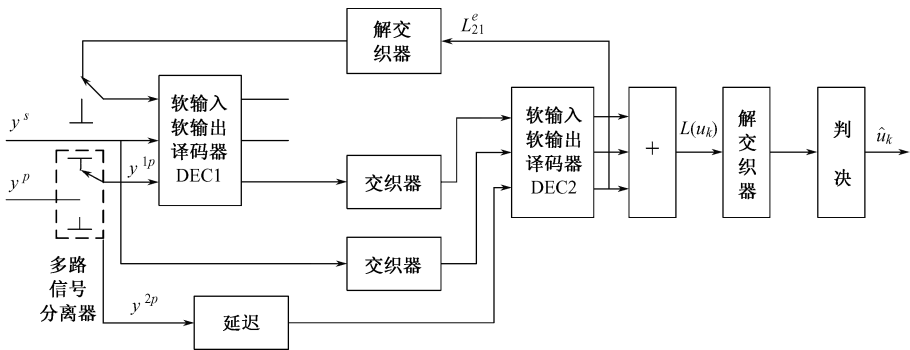


图 9-20 Turbo 码译码器基本结构图

假定 Turbo 码译码器的接收序列为  $\mathbf{y}=(\mathbf{y}^s, \mathbf{y}^p)$ ，冗余信息  $\mathbf{y}^p$  经复用后，分别送给 DEC1 和 DEC2。于是两个 SISO 译码器的输入序列分别为

$$\text{DEC1: } \mathbf{y}_1=(\mathbf{y}^s, \mathbf{y}^{1p}) \qquad \text{DEC2: } \mathbf{y}_2=(\mathbf{y}^s, \mathbf{y}^{2p}) \qquad (9-46)$$

为了使译码后的比特错误概率最小，根据最大后验概率译码准则，Turbo 码译码器的最



佳译码策略是：根据接收序列  $\mathbf{y}$  计算后验概率  $P(u_k)=P(u_k|y_1, y_2)$ 。但由于计算复杂度随码长的增加而增加，这种最佳策略会变得不可实现。因此，在 Turbo 码的译码方案中，巧妙地采用了一种次优译码规则，将  $y_1$  和  $y_2$  分开考虑，由 2 个分量码译码器分别计算后验概率  $P(u_k)=P(u_k|y_1, L_1^e)$  和  $P(u_k)=P(u_k|y_2, L_2^e)$ ，然后通过 DEC1 和 DEC2 之间的多次迭代，使它们收敛于  $P(u_k|y_1, y_2)$ ，从而达到接近 Shannon 极限的性能。这里  $L_1^e$  和  $L_2^e$  为附加信息，其中， $L_2^e$  由 DEC2 提供，在 DEC1 中用做先验信息； $L_1^e$  由 DEC1 提供，在 DEC2 中用做先验信息。

因篇幅所限，更多的有关 Turbo 码方面的问题，不再作介绍，感兴趣读者可以参考相关文献。

## 思考与练习

- 9-1 信道编码与信源编码有什么不同？纠错码能够检错或纠错的根本原因是什么？
- 9-2 差错控制的基本工作方式有哪几种？各有什么特点？
- 9-3 分组码的检（纠）错能力与最小码距有什么关系？
- 9-4 系统分组码的监督矩阵、生成矩阵各有什么特点？相互之间有什么关系？
- 9-5 简述分组码和卷积码的区别。
- 9-6 已知  $(7, 3)$  循环码的生成多项式为  $g(x)=x^4+x^2+x+1$ 。若  $m(x)$  分别为  $x^2$  和 1，求其系统码的码字。

9-7 GSM 系统中全业务信道使用的卷积码生成多项式如下：

$$g_1(x)=1+x^3+x^4$$

$$g_2(x)=1+x+x^3+x^4$$

画出编码器框图。

9-8 一个编码效率  $R=1/2$  的卷积码编码器如图 9-21 所示，求由信息序列 10111...产生的编码器输出。

9-9 效率为  $1/2$ 、约束长度为 3 的卷积码的网格图如图 9-22 所示，如果传输的是全 0 序列，接收到的序列是 100010000...，利用维特比译码算法，计算译码序列。

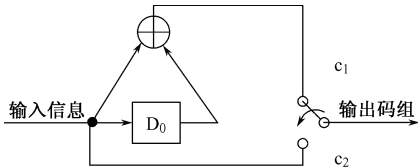


图 9-21 卷积码编码器

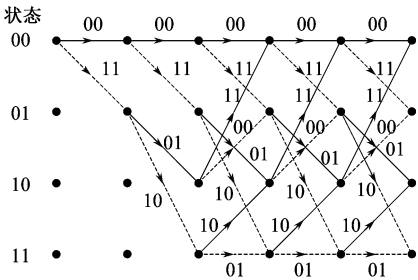


图 9-22 卷积码网络图

9-10 试解释 TCM 的原理，画出 8PSK 星座图的集合分割，并计算最小欧氏距离。

## 第 10 章 通信加密解密与密码分析

密码学是研究信息系统安全保密的科学，它包含两个分支，即密码编码学和密码分析学。所谓的密码编码学就是对信息进行编码实现隐蔽信息，而密码分析学就是研究分析破译密码的学问，两者相互对立又互相促进地向前发展。随着以计算机技术和网络通信技术为代表的信息技术的不断发展和迅速普及，通信保密问题日益突出。

### 10.1 密码学基本概念

加密，即密码编码，是对信息进行编码实现隐蔽信息的一门学问。解密，即脱密，是加密的逆过程，是掌握密钥和密码体制（算法）的人们从密文恢复出原文的过程。密码分析，是在不知道密钥和密码体制的情况下，企图通过其他途径发现明文或密钥的过程，是研究分析破译密码的学问。解密和加密一样，都使用确定的数学工具。解密者掌握了密钥和密码体制（算法），密码分析者则没有掌握密钥和密码体制（算法）；密码分析除了依靠数学、工程背景、语言学等知识外，还要靠经验、统计、测试、眼力、直觉判断能力等，有时也要靠点运气。

#### 10.1.1 密码学发展历史

密码学在发展的过程中，经历了古典密码学、近代密码学和现代密码学 3 个阶段。随着计算机技术的发展，密码算法越来越复杂，密码体制的设计正在由实际不可破向理论上完全保密的方向发展。

##### 1. 古典密码学

第一个阶段是从古代到 19 世纪末，长达数千年，这个时期生产力低下，产生的许多密码体制都是可用纸笔或者简单的器械实现加解密的。我们称这个时期产生出的密码体制为“古典密码体制”。这个阶段所产生出来的所有密码体制几乎已全部被破译了，但是了解它们的设计原理和破译分析方法对现代电子密码体制的设计和破译都是有一定意义的。

下面介绍几种古典密码体制。

（1）凯撒密码。意大利凯撒大帝在给其前线将军写信时，使用一种单字符单表代换密码，将 26 个字母按自然顺序结成一个圆圈，写信时，他将每个明文字母都用该字母后面的第三个字母来代替。如图 10-1 所示。



图 10-1 自然顺序结成的字母圆圈

对于明文“computer”，取  $k=3$ ，加密  $E_k(\text{computer}) = \text{frosxwhu}$ 。  
这种单字符单表代换密码体制，对相同明文字母有最大  $K=26$  种可能的密文。  
(2) 单字符多表代换密码。例如，对于明文字母“abcdefghijklmnopqrstuvwxyz”，有单字符多表代换密表：

- 密表 1: xyztheowrflndabcgijkmpqsuv
- 密表 2: ghijklmnopqrstuvwxyzabcdef
- 密表 3: vwxyzinformatsecuybdghjklpq

若要发送的明文为“comupter network”，并将每 3 个字符分为一组，第 1 组中的字符按密表 1 变换（计为  $\pi_1$  变换），第 2 组中的字符按密表 2 变换（计为  $\pi_2$  变换），第 3 组中的字符按密表 3 变换（计为  $\pi_3$  变换），周期为 3。则

$$\begin{aligned} E_k(\text{computer network}) &= \pi_1(\text{com}) \pi_2(\text{put}) \pi_3(\text{ern}) \pi_1(\text{etw}) \pi_2(\text{ork}) \\ &= \text{zbdvazibehkquxq} \end{aligned} \tag{10-1}$$

2. 近代密码学

密码学第二个阶段从 20 世纪初到 20 世纪 50 年代末，这半个世纪期间，由于莫尔斯发明了电报，为了适应电报通信，密码设计者设计出了一些采用复杂的机械和电动机械设备实现加解密的体制，我们称这个时期产生出的密码体制为“近代密码体制”。

近代密码体制是用机械或电动机械实现的，其最基本的东西是转轮机。一个转轮如果固定使用，只能起到单表代替的作用。只有使轮子转动起来才能起到多表代替的作用。

假如在输入板和输出板之间安放 3 个字母转轮（3 个加密表），当第一个字母转轮转动一圈时，第二个字母转轮才转动一格，当第二个字母转轮转动一圈时，第三个字母转轮才转动一格，由于每个字母转轮都有 26 个不同位置，故 3 个转轮结合在一起就有  $26 \times 26 \times 26 = 17\,576$  个不同位置，就可以产生一个周期为 17576 的多表代替加密体制。

同样，输入板和输出板之间安放多个字母转轮（多个加密表），当第一个字母转轮转动一圈时，第二个字母可转轮转动两格，当第二个字母转轮转动一圈时，第三个字母可转轮转动三格，……，故多个转轮结合在一起就有更多不同位置，即采用复杂的机械和电动机械转轮机设备，可实现更复杂的密文。

以下介绍几种“近代密码体制”的机械和电动机械设备。

(1) Kryha 密码机，如图 10-2 所示，大约在 1926 年由 Alexander vo Kryha 发明。这是一个多表加密设备，密钥长度为 442，周期固定。是一个由数量不等的齿轮引导密文轮不规则运动。

(2) M-209，如图 10-3 所示，是哈格林对 C-36 改进后的产品，由 Smith-Corna 负责为美国陆军生产。它的密码周期达到了 101, 105, 950。



图 10-2 Kryha 密码机



图 10-3 M-209

总之，若有 26 个转轮（26 个加密表），每个最大移位  $K=26$  时，密钥量达  $26^{26}$ ，安全吗？对这些密码体制及方法，现在已被证明是不保密的，但是要想破译它们往往需要很大的计算量。

### 3. 现代密码学

第三个阶段是从香农（Shannon）于 1949 年发表的划时代论文（Communication Theory of Secrecy System）“保密体制的通信理论”开始的。这篇论文证明了密码编码学有着坚实的数学基础。微电子技术的发展使电子密码走上历史舞台，催生了“现代密码体制”。特别是本世纪 70 年代中期，数据加密标准（Data Encryption Standard, DES）密码算法的公开发表，以及公开密钥思想的提出，更是促进了当代密码学的蓬勃发展。大规模集成电路技术和计算机技术的迅速发展，不仅给密码设计者带来好处、便利和成功，也给密码分析者带来好处、便利和成功。

现代密码体制强调保密性只建立在密钥安全上，加密密钥可以公开，解密密钥保密，并从加密密钥中推算出解密密钥是非常非常困难的事，即现代密码体制安全性建立在难解的数学问题上。

### 10.1.2 密码体制模型

密码体制一般模型如图 10-4 所示。其中，明文是未被加密的信息，而“消息源”中所有可能的明文组合在一起就叫做“明文空间”（M），明文空间中的所有可能形式都是人们可以理解的或者借助于可公开得到的工具就可理解的。

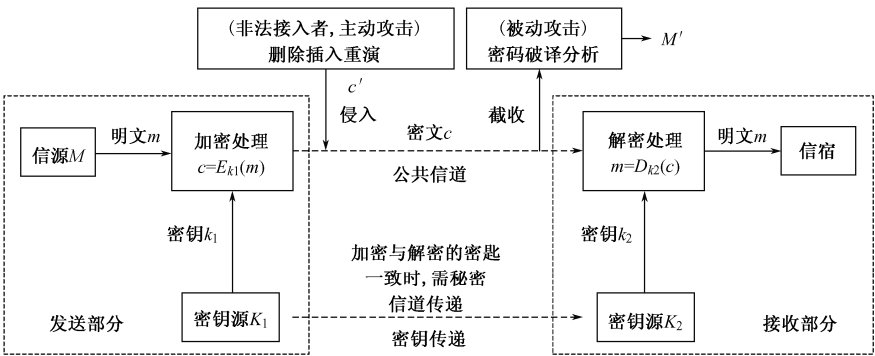


图 10-4 密码体制的一般模型

同样，密文就是被加密的消息，所有可能的密文亦可组合在一起叫做“密文空间”（C）。密文空间中的所有可能形式应当是人们不可以直接理解的。

密钥就是控制加密和解密算法操作的数据处理，分别称为加密密钥和解密密钥，那么，所有的可能密钥也组成一个“密钥空间”（K）。

一般说来，这 3 个空间的元素数量都是有限的，分别记为  $|M|$ 、 $|C|$  和  $|K|$ 。假定有甲乙两个人需要进行相互间的保密通信，并假定其中甲方为发信者，乙方为收信者。甲乙两方欲通过对称体制进行保密通信，就需按下述步骤进行操作：

- （1）甲、乙双方商定好从密钥空间  $K$  中选定一个双方共知而他人不知的密钥  $k$ ；
- （2）甲方从明文空间选取要发送给乙方的明文  $m$ ，用确定的加密变换将明文  $m$  转变成

密文  $c$ ，即  $c = E_k(m)$ ；

(3) 将密文  $c$  用通常的通信工具传送给乙方；

(4) 乙方将收到的密文  $c$  用密钥  $k$  确定的解密变换还原成明文  $m$ ，即  $m = D_k(c)$ 。

采用加密的方法可以隐蔽和保护需要保密的消息，使未授权者不能提取信息。密钥可将明文变换成另一种隐蔽的形式，称为密文或密报。这种变换过程称为加密。其逆过程，即由密文恢复出原文的过程称为解密。对明文进行加密操作的人员称为密码员。密码员对明文进行加密时所采用的一组规则称为加密算法。传送消息的预定对象称为接收者，他对密文进行解密时所采用的一组规则称为解密算法。加密和解密算法的操作通常都是在—组密钥控制下进行的，分别称为加密密钥和解密密钥。

密钥是密码的“关键”、“要害”、“秘诀”，确保密钥本身的安全性是非常重要的，是密码体制安全保密的关键，它的产生和管理是密码学中的重要研究课题。

### 10.1.3 密码体制的分类

从不同的角度对密码体制有不同的分类。按照加密解密使用密钥是否相同，可以把密码体制分为单钥密码体制（或对称密码体制）和公钥密码体制（或非对称密码体制）；按照加解密时是否对明文进行分组，可以分为分组密码体制和序列密码体制（或流密码体制）；按照将明文转变成密文的操作类型，可分为单表代换与多表代换等。下面重点讨论单钥密码和公钥密码体制。

#### 1. 单钥密码（对称密码体制）

单钥密码的特点是无论加密还是解密都使用同一个密钥，因此，此密码体制的安全性就是密钥的安全，如果密钥泄露，则此密码系统便被攻破。

如何产生满足保密要求的密钥是这类体制设计和实现的主要课题，另一个重要问题是如何将密钥安全可靠地分配给通信对方。在网络通信条件下就更为复杂，包括密钥产生、分配、存储、销毁等多方面的问题，统称为密钥管理（Key Management）。这是影响系统安全的关键因素，即使密码算法再好，若密钥管理问题处理不好，就很难保证系统的安全保密。

单钥加密的古典算法主要有单表代换和多表代换，最有影响的单钥密码是1977年美国国家标准局颁布的数据加密标准（Data Encryption Standard, DES）。

单钥密码的优点是：安全性较高，加解密速度快，适合大量数据信息的加密。

缺点是：

- (1) 随着网络规模的扩大，密钥的管理成为一个难点；
- (2) 无法解决消息确认问题；
- (3) 缺乏自动检测密钥泄露的能力。

#### 2. 公钥密码（非对称密码体制）

公开加密密钥（public-key）体制也是非对称密码体制。在公钥体制下，加密密钥与解密密钥不同。此时不需要安全信道来传送密钥，而只需利用本地密钥发生器产生解密密钥并以此来控制解密操作。1976年W.Diffie和M.E.Hellman首先在论文“密码学的新方向”中提出了公开密钥体制的概念。

### 1) 核心思想

公钥密码体制的核心思想是加密和解密使用不同的密钥，加密密钥可以被任何人知道（公钥），保密解密密钥（私钥），并要求不得从加密密钥中推算出解密密钥，或者说，从加密密钥中推算出解密密钥是非常非常困难的事，即不存在有效的算法。这是密码学几千年历史中最有革命性的结果。

这样的公钥系统，加密密钥可以像电话号码本一样公开，任何人都可以查到该用户的公开密钥，用它来加密给该用户发送密文，该密文只有该用户拥有解密密钥，任何第二者都无法获得解密密钥和明文。另外，可以实现多个用户加密的消息只能由一个用户解读，或只能由一个用户加密消息而使多个用户可以解读，前者可用于公共网络中实现保密通信，而后者可用于密钥安全传输和安全认证。认证系统中对消息进行数字签字，实现安全鉴别，但速度慢。

### 2) 优缺点

优点：将加密和解密能力分开，通信前不需要密钥的分发，不存在密钥安全传输问题；通信实体间可实现一对多或多对一通信。

缺点：加解密速度慢。

### 3) 公钥密码经典算法举例

第一个比较完善的公开密钥经典算法是由 Rivest、Shamir 和 Adleman 三人在 1977 发明的 RSA 算法。

RSA 为分组加密算法，其安全性是基于大素数相乘和因子分解，可以被看成一个单向函数，其相乘容易，分解很难，即加密函数  $E_k(x)=x^e \bmod n$  是一个单向函数，而解密的陷门是知  $\Phi(n)=(p-1)(q-1)$ ，分解  $n=pq$ ，所以对攻击的人来说求逆计算不可行。

设明文和密文在  $0 \sim (n-1)$  之间， $n$  是一个正整数，对明文块  $M$  和密文块  $C$ ，RSA 加密和解密的形式如下：

$$C = M^e \bmod n \tag{10-2}$$

$$M = C^d \bmod n = (M^e)^d \bmod n = M^{ed} \bmod n \tag{10-3}$$

发送方和接收方都必须知道  $n$  和  $e$  的数值，而只有接收方知道  $d$  的数值。这就是使用公钥  $K_U=\{e, n\}$  和私钥  $K_R=\{d, n\}$  的一种公钥加密算法。要是此算法满足公钥加密，必须符合以下要求：

- (1) 能够发现  $e$ 、 $d$ 、 $n$  的值，其中对于所有的  $M < n$ ， $M^{ed} = M \bmod n$ ；
- (2) 对所有  $M < n$  的数值， $M^e$  和  $C$  的计算相对简单；
- (3) 给定  $e$  和  $n$ ，不能得出  $d$ 。

前两个要求很容易实现，第三个要求对于较大的  $e$  和  $n$  才能实现。

开始先选择两个素数  $p$  和  $q$ ，计算出它们的积  $n$ ，这是加密解密的模数；接着需要知道  $n$  的欧拉函数  $\Phi(n)$ ；然后选择整数  $e$ ，它是  $\Phi(n)$  的相对质数；最后计算  $d$  作为  $e$  的乘法逆元，即模  $\Phi(n)$ ，具体如图 10-5 所示。

|  |                |
|--|----------------|
| RSA密钥生成  |                |
| 1. 选择 $p, q$ ;                                       | ( $p, q$ 都是素数) |
| 2. 计算 $n=pq$ ;                                       |                |
| 3. 计算 $\Phi(n)=(p-1)(q-1)$ ;                         |                |
| 4. 选择整数 $e$ , $\gcd(\Phi(n), e)=1$ 且 $1<e<\Phi(n)$ ; |                |
| 5. 计算 $d$ , $de \equiv 1 \pmod{\Phi(n)}$ ;           |                |
| 公钥: $K_U=\{e, n\}$ ; 私钥: $K_R=\{d, n\}$ 。            |                |
| 加密   |                |
| 明文: $M < n$  |                |
| 密文: $C = M^e \pmod{n}$                               |                |
| 解密   |                |
| 密文: $C$  |                |
| 明文: $M \equiv C^d \pmod{n}$                          |                |

注： $\Phi(n)$ 为欧拉函数，它是比 $n$ 小的正整数，是 $n$ 的相对素数；  
 $\gcd(\Phi(n), e)$ 表示 $\Phi(n)$ 和 $e$ 的最大公约数；  
 $a \bmod n$ 定义成 $a$ 除以 $n$ 的余数。

图 10-5 RSA 算法

密码分析者攻击 RSA 体制的关键点在于如何分解 $n$ 。若分解成功使 $n=pq$ ，则可以算出 $\Phi(n)=(p-1)(q-1)$ ，然后由公开的 $e$ ，解出秘密的 $d$ 。若使 RSA 安全， $p$ 与 $q$ 必为足够大的素数，使分析者没有办法在较短的时间内将 $n$ 分解出来。

假定用户 A 已经发布了它的公钥，用户 B 想要向 A 发送消息 M，本例中假定生成的密钥如下：

- 设  $p=7, q=17, n=7 \times 17=119$ ;
- $\Phi(n)=(7-1)(17-1)=96$ ;
- 选择  $e=5, \gcd(5, 96)=1$ ;
- 计算  $d$ , 使 $(d \times e) \bmod 96 \equiv 1$ ，正确数值为  $d=77$ ，因为  $77 \times 5=385=4 \times 96+1$ ;
- 公钥 $\{5, 119\}$ ，私钥 $\{77, 119\}$ 。

设明文  $M=19$ ，则加密： $(19)^5 \bmod 119 = 66$ ；解密： $(66)^{77} \bmod 119 = 19$ 。

### 10.1.4 Shannon 的保密通信理论

香农于 1948 年发表了《通信的数学理论》，这篇论文产生了信息论。他又于 1949 年发表了《保密体制的通信理论》，这是密码学发展历史上的一篇划时代的论文。信息论（亦有人称之为“通信的数学理论”）是研究信息率、信道容量、噪声以及其他影响信息传输因素的一种数学理论。将信息论应用于密码学，产生了许多新的概念，其中主要有：多余度和唯一解距离，理论保密和实际保密，完全保密体制和理想保密体制，扩散原理和混乱原理，等等。

#### 1. 多余度和唯一解距离

多余度的概念，从字面上来看，就是“多余的程度”，指的是在通信过程中，所传送的符号比信息所需要的符号要多。这多余的部分在检错和纠错技术中是很有用的。

在信息的传输和处理过程中，出错是难免的，为了检测错误和纠正错误，还要另外添加一些数据，如作奇偶校验用。这样做，对通信的可靠性是大有好处的。

但是，在密码学中这种多余度的存在就不是好事了。比如说，英语中的 the、of、and、to、a、in、that、it、is 和 I 这 10 个词汇，单独出现时是没有意义的，但它们却是英语中最常用的单词，任何一篇文章中都有四分之一的篇幅被这 10 个词汇占据。再比如说，对汉语中字词出现的频率进行统计的结果表明，“的、了、一、是、不、我、在、个、有、着”这 10 个字在汉语中出现的频数最大，合在一起竟达 17.675 75%。英文单字母的相对频率表如图 10-6 所示，英文单字母使用频率分类如表 10-1 所示。

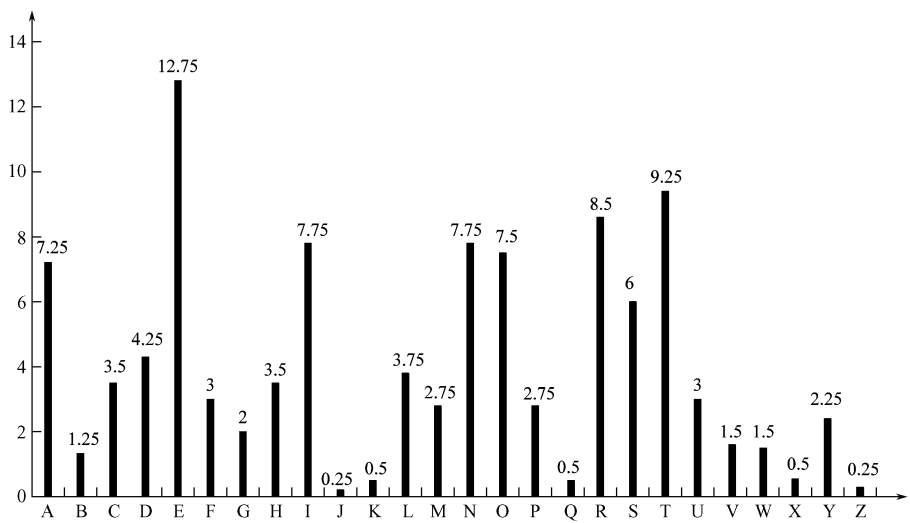


图 10-6 英文单字母的相对频率表

表 10-1 英文单字母使用频率分类

| 分 类 | 使用频率分类字母集                         | 每个字母约占比例    |
|-----|-----------------------------------|-------------|
| 1   | 极高使用频率字母集：e                       | 12 %        |
| 2   | 次高使用频率字母集：t, a, o, l, n, s, h, r  | 6 %~9 %     |
| 3   | 中使用频率字母集：d, i                     | 4 %         |
| 4   | 低使用频率分类字母集：c, u, m, w, g, y, p, b | 1.5 %~2.3 % |
| 5   | 次低使用频率分类字母集：v, k, j, x, q, z      | 1%          |

由于这种情况的存在，密码分析者在进行破译分析时就有了可利用的东西。密文中出现频数最多的那些字母、字和词就很可能是在明文中出现频数最多的那些字母、字和词。对密码分析者有利，对密码设计者就不是好事。香农指出：多余度为密码分析奠定了基础。这是香农对密码学的巨大贡献。多余度越低，破译分析的困难性越大，这对保密却是有利的。

怎样来衡量唯密文破译分析的困难性？香农定义了一个称为“唯一解距离”（又称“唯一解码量”）的量  $v_0$ 。

当密码分析者所截获的密文字符数大于  $v_0$  时，从原则上讲可以破译该密码系统。香农利用信息论描述了被截获的密文量和成功攻击的可能性之间的关系，他列出了一个比较复杂的数学公式来计算唯一解距离。这个数学公式虽然因密码体制的不同而有所不同，但总是把多余度这个量作为分母，即可以表示为  $v_0 \approx H(K)/\delta_l$ ，其中  $H(K)$  为密钥熵， $\delta_l$  为  $l$  长二进制明文序列的多余度。



2. 理论（完全）保密和实际保密

香农在对理论保密性的研究中，得出了完全保密和理想保密两种体制，它们都是理论上不可破译的好体制，无论密码分析者有多少时间和人力，也无论他能截获多大密文量，他都破译不了这两种密码体制（请注意，“破译不了”是对“唯密文攻击（只有密文可供破译）”而言的）。

1) 安全保密

对于任意给定的密码体制，截取任意长度的密文，用每个密钥进行解密所获得的有意义的消息数量越多，那么密码分析者判定实际发送的是哪个消息就越难。最理想的是，用所有的密钥解密时，解密的结果都有意义，这样就不会给密码分析者任何一点信息，这就是完全密码的概念。

设明文空间  $M=\{m_1, m_2, \cdots, m_r\}$ ，密文空间  $C=\{c_1, c_2, \cdots, c_s\}$ ，密钥空间  $K=\{k_1, k_2, \cdots, k_r\}$ ，并令  $p(m_i)$  是  $m_i$  的先验概率。

若截获密文  $c_j$ ，求得在已知  $c_j$  的条件下， $m_i$  被发送的后验概率  $p_j(m_i)$ ，如果对每个明文  $m_i$  和每个密文  $c_j$ ，都有

$$p(m_i)=p_j(m_i)$$
 (10-4)

那么，截获密文  $c_j$  对于原文毫无帮助，这时的加密系统就称为理论上完全保密。

完全保密体制中，明文数、密钥数和密文数都相等，将每个明文变换成每个密文都恰好有一个密钥，所有的密钥都是等可能的。完全保密没有给密码分析者任何额外的可用于破译的信息，因此密码分析者无法破译这种体制。但是为了确保完全保密的代价太大，那就是必须使不同的密钥数至少要和可能的明文数一样多。在消息空间较小时还可以，当消息空间较大时，密钥管理就成了大问题。

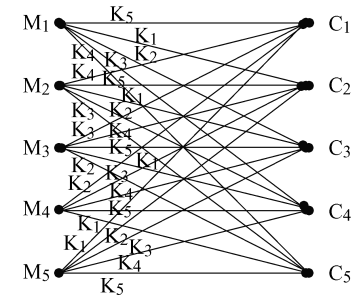


图 10-7 完全保密体制之例

虽则如此，完全保密的思想至今还在指导着当代的密码设计者。

图 10-7 所示的完全保密体制例子是香农首先给出的，在此密码体制中，明文为  $M_1\sim M_5$ ，密钥为  $K_1\sim K_5$ ，密文为  $C_1\sim C_5$ ，它们的数量均是 5。

2) 理想保密

当一个密码体制的唯一解距离趋向于无穷大时，该密码体制就是理想保密体制。

此时，无论密码分析者截获了多少密文消息都无助于破译该密码体制。 $v_0$  趋向于无穷大，意味着语言的多余度趋向于零。如果要消除语言中的全部多余度，事实上也是不可能的。看来这种体制也是不实用的。但是，这可以提示我们，在设计密码体制（算法）时，应尽量缩小多余度。加密之前先对明文消息进行信源压缩编码就可达到此目的。

3) 实际保密

考虑时间、设备、资金，要求在掩盖时间内保密，即在有效时间内实际所需的、计算资源是不可实现的、或计算安全的，称为实际保密。

#### 4) 实际保密对密码体系的要求

为了保护信息的保密性，抗击密码分析，保密系统应当满足下述要求：

(1) 系统即使达不到理论上是不可破的，也应当是实际上不可破的，就是说，从截获的密文或某些已知明文密文对，要决定密钥或任意明文在计算上是不可行的；

(2) 系统的保密性不依赖于对加密体制或算法的保密，而依赖于密钥，这是著名的 Kerckhoff 原则；

(3) 加密和解密算法适用于所有密钥空间中的元素；

(4) 系统便于实现和使用方便。

### 3. 扩散原理和混乱原理

为挫败统计分析，香农提出了两种实用的方法：扩散法和混乱法。

所谓扩散，是将每一位明文字符的影响尽可能地散布到较多输出密文字符中去，以便隐藏明文字符的统计特性。对密钥的处理也应如此。

所谓混乱，是使密文和明文、密钥之间的关系复杂化，给任何依靠统计分析来揭示密钥特征的工作制造更多困难。

扩散和混乱的结果，就使得密码分析需要截获非常多的密文数据才能试图用统计分析方法进行破译工作。但是也带来一个缺点：在接收端进行解密时，会由于传输中的一个错误而引起许多错误出现。通常称这种现象为“错误扩散”。产生这种现象的原因就是，由于“每位明文数字的影响尽可能地散布到多个输出密文数字中”，那么，这“输出密文数字”中的某一个如果在传输过程中出错，就会影响多位明文数字在接收端的恢复。

## 10.2 解密与密码分析

在密码学发展的各个阶段，密码编码学和密码分析学在相互竞争中都得到了发展，一方编制了不少各种各样的密码体制，总认为每种密码体制的保密性多么强；另一方则凭借自己的经验，依赖客观世界中存在的事实，针对每种密码体制的特点进行分析破译。

### 10.2.1 解密和密码分析

#### 1. 解密

解密（或称“脱密”）是加密的逆过程，是正当掌握密钥和密码体制（算法）的人们从密文恢复出明文的过程。

#### 2. 密码分析

有一部分人出于某种目的（或是为了获取情报，或是为了好奇、感兴趣），虽然他们没有掌握密钥或密码体制（算法），也企图破译通过各种门路所获得的密文，设法读懂密文。这就是“密码分析”，又称“密码破译”。

那么，密码分析（攻击）的种类有以下几种：

- （1）唯密文攻击——只有密文可供破译；
- （2）已知明文攻击——已知一些明文及其对应密文；
- （3）选择明文攻击——密码分析者能够选择一些明文，并构造出这些明文所对应的密文来；
- （4）选择密文攻击——密码分析者能够选择一些密文，并构造出这些密文所对应的明文来。

上述几种攻击的强度依次增大，唯密文攻击最弱。破译密码的基础是依赖于明文中的多余度，攻击目的是确定密码体制中的当用密钥。抗不住唯密文攻击的密码体制（算法）就是可破译，不能使用的。

注意，“解密（脱密）”和“密码分析”二者虽然都是设法将密文还原成明文，但是二者的含义和所使用的方法是不同的。首先，解密者掌握了密钥和密码体制（算法）；而密码分析者则没有掌握密钥和密码体制（算法）。其次，解密和加密一样，都使用确定的数学工具；而密码分析除了依靠数学、工程背景、语言学等知识外，还要靠经验、统计、测试、眼力、直觉判断能力等，有时也要靠点运气。

## 10.2.2 密码分析的方法

密码分析者通常采用科学家们在解决自然科学问题时常用的研究步骤，即：“分析—假设—推断—证实（或否定）”的4步作业方法，基本数学工具是统计分析、数学演绎和归纳。

密码分析方法有穷举破译法和分析破译法两类。

### 1. 穷举法

穷举法是对截收的密文依次用各种可解的密钥试译，直到得到有意义的明文；或在不变的密钥下，对所有可能的明文加密直到得到与截获密文一致为止，此法又被称为完全试凑法。只要有足够多的计算时间和存储容量，原则上穷举法总是可以成功的。但实际上，任何一种能保障安全要求的实用密码都会设计得使这一方法在实际上是不可行的。

为了减少搜索计算量，可以采用较有效的改进试凑法。它将密钥空间划分成几个等可能的子集，对密钥可能落入哪个子集进行判断，在确定了正确密钥所在的子集后，就对该子集再进行类似的划分并检验正确密钥所在的子集，依次类推，最终即可判断出所用的正确密钥了。

### 2. 分析破译法

分析破译法有确定性和统计性两类。确定性分析法是利用一个或几个已知量（比如已知密文或明文—密文对）用数学关系式表示出已知量与所求未知量（如密钥等）的关系。寻求这种数学关系是确定性分析法的关键步骤。统计分析法是利用明文已知的统计规律进行破译的方法。密码破译者对截收的密文进行统计分析，总结出其间的统计规律，并与明文的统计规律进行对照比较，从中提取出明文和密文的对应和变换关系。

一种密码分析的复杂度可以分为两部分：数据复杂度和处理复杂度。数据复杂度是实施该攻击所需输入的数据量；而处理复杂度是处理这些数据所需的数据量。这两部分的主要部

分通常被用来刻画该攻击的复杂度。

## 10.3 通信系统与网络保密技术

随着以计算机技术和网络通信技术为代表的信息技术的不断发展和迅速普及，通信保密问题日益突出。信息安全问题已经成为阻碍经济持续稳定发展和威胁国家安全的一个重要问题。众所周知，密码是信息安全的核心，设计具有自主知识产权的新型高性能的密码体制是目前最亟待解决的重要问题。本节主要讲述语音图像保密技术及通信网的保密方式等。

### 10.3.1 语音保密通信技术

语音是人们获取信息的重要手段，语音通信是现代通信中最常用、最方便的通信手段之一，所以研究语音通信保密技术具有很大的意义。

现代语音通信分模拟语音信号通信与数字语音信号通信，由于模拟语音为连续变化量，非前面讨论的数字、数据或字符加密问题，所以首先我们介绍几种模拟语音通信加密技术，然后再介绍数字语音信号加密原理，最后介绍基于混沌的语音掩盖加密。

#### 1. 模拟语音信号倒频法

模拟语音信号频率倒置是最古老的加密方法之一，它利用倒置器将所送的语音信号频率在被传送之前进行倒置处理。例如，对语音信号来说，假如取语音信号的最高频率为 3 000 Hz，则可将 300~3 000 Hz 的语音频谱的高频成分移到低频，低频成分移到高频，如图 10-8 所示。

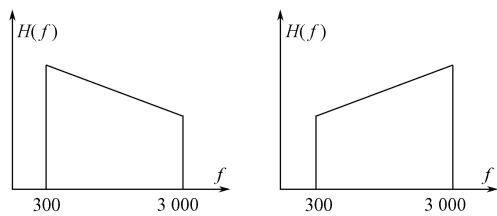


图 10-8 模拟语音信号倒频法原理图

这种加密技术的优点是技术简单，设备经济，使用方便。其缺点是保密性差，只能防止临时窃听，防止那些毫无专门训练的人窃听，如窃听者长时间反复对录音带进行放音收听，或者窃听者是专门经过听力训练的人，则他们可听明白绝大部分倒频的语音。所以，这种简单的倒频技术单独使用的价值不高。

#### 2. 模拟语音信号频段分割移动法

这种加密方法是通过频段分割器将信号的频谱分成若干频段。例如，对语音信号来说，可将 2 700 Hz 的音频信号分成 4~5 个单独的频段，每个频段宽为 500 Hz 左右；然后再将这些频段改变次序进行传输。例如，频段 1 的信号可移在频段 3，频段 2 的信号可移在频段 2 或频段 5 等，如图 10-9 所示。到了接收端再将这些频段的信号移到原来位置即可实现解密。这种加密方法较前一种加密方法保密性稍好一些，设备也不复杂。但是保密程度仍然较低。

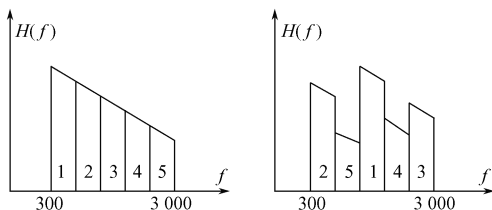


图 10-9 模拟语音信号频段分割移动法原理

### 3. 模拟语音信号频段分割移动及倒置法

这是一种将分割移动与倒置相结合的加密方法，它不仅将各段的信号移动位置，而且还倒置它本身的高低频端的关系，如图 10-10 所示。

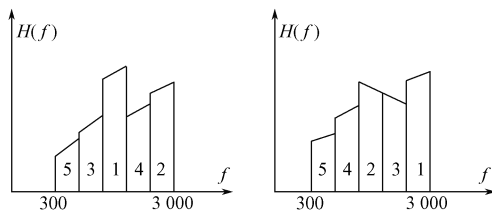


图 10-10 模拟语音信号频段分割移动及倒置法

### 4. 模拟语音信号滚码频段分割倒置法

这种加密方法就是以每秒 2~4 次的变换速度来重新排列其频段的顺序，图 10-11 所示为一种每秒变换 4 次的情况。图 10-11 中 A 帧为一种 4 频段的排列，而 B、C、D 各帧为不同频段的排列。这种排列序列将周期性的出现。在接收端以相同的变换速率进行反置，即可实现译码。

滚码频段分割倒置法较之上述普通位静态频段分割倒置法保密性要好，而频段分割越细，每秒钟变换帧数越多，保密性越好。但是分割得太细时，恢复的语音质量将下降。

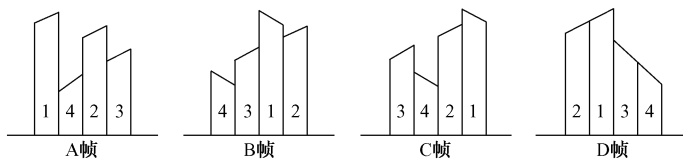


图 10-11 模拟语音信号滚码频段分割倒置法

### 5. 数字语音信号加密原理

数字语音信号加密原理，如图 10-12 所示，首先采用波形编码或参量编码将语音信号数字化，然后对数字语音信号加密，这样就可以利用所有数字（数据或字符）信号加密技术，完成加密。

### 6. 基于混沌的语音掩盖加密

近几年来，人们将混沌信号和语音加密技术相结合，使得混沌技术在语音保密通信中的应用研究更具有现实意义。

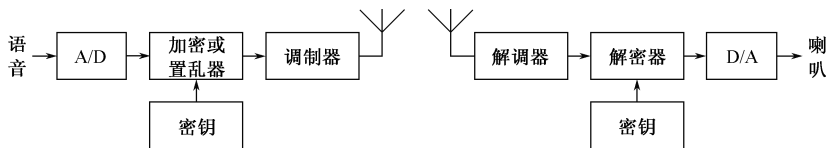


图 10-12 数字语音信号加密原理

混沌（Chaos）是一种貌似无规则的运动，指在确定性非线性系统中，不需附加任何随机因素也可以出现类似随机的行为（内在随机性）。混沌系统的最大特点就在于系统的演化对初始条件极端敏感。因此，从长期意义上讲，系统的未来行为是不可预测的。

混沌系统都具有如下基本特性：确定性、有界性、对初始条件的敏感性、拓扑传递性和混合性、宽带性、快速衰减的自相关性、长期不可预测性和伪随机性。

混沌系统所具有的这些基本特性恰好能够满足保密通信及密码学的基本要求：

- （1）混沌动力学方程的确定性保证了通信双方在收发过程或加解密过程中的可靠性；
- （2）混沌轨道的发散特性及对初始条件的敏感性正好满足 Shannon 提出的密码系统设计的第一个基本原则——扩散原则；
- （3）混沌吸引拓扑传递性与混合性，以及对系统参数的敏感性正好满足 Shannon 提出的密码系统设计的第二个基本原则——混乱原则；
- （4）混沌输出信号的宽带功率谱和快速衰减的自相关特性是对抗频谱分析和相关分析的有力保障；
- （5）混沌行为的长期不可预测性是混沌保密通信安全性的根本保障等。

基于混沌的语音掩盖加密方法。它具有抗破译能力强，对噪声具有免疫力，被还原的语音信号没有失真等优点。用这种方法进行语音加密和解密都将取得了非常理想的效果。

这里对其具体的基于混沌的语音掩盖加密方法不作具体介绍。

### 10.3.2 图像保密通信技术

图像保密通信指将一幅原始图像经过某个加密算子的作用，变为可传输的密文图像，接收方将密文图像经过解密算子的作用，还原为原始图像。在传输过程中，即使密文图像被窃取，没有密钥很难或根本无法完成解密。

图像加密算法往往可以写成如下变换式

$$Y=E_k(X) \quad (10-5)$$

其中，明文  $X$ 、密文  $Y$  和密钥  $k$  分别是有限长度的字符序列，它们的长度  $L_X$ 、 $L_Y$  和  $L_k$  可以互不相等，该方程强调一个密文  $Y$  与一个明文  $X$  对应，函数由密钥的值来决定； $E_k$  称为加密算法。

对于图像进行保密通信，现在普遍采用比较先进的图像数据混沌加密技术（算法），加密解密原理如图 10-13 所示。

基于混沌的图像掩盖加密技术就是利用混沌动力学系统实现对图像数据块的某种变换，变换后使得图像与原始图像存在明显的视觉差异，从而实现对原始图像加密的目的。

按照变换对象的不同，混沌图像加密可以分为图像像素值变换、像素值坐标变换以及这两种变换方式的混合的 3 种情况。按照混沌系统的不同，可以分为一维、二维、多维时空混

沌系统加密算法。按照传统密码学的分类，可以分为混沌分组密码、流密码。

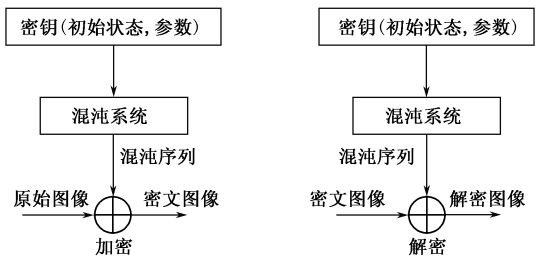


图 10-13 加密解密算法模型

下面我们介绍一种具体图像加密技术，基于双一维混沌映射的图像分组加密算法的。这种方法利用混沌系统产生混沌序列，进行适当的处理，将得到的序列与原始图像进行作用，得到密文图像。密钥取混沌系统的初始值和参数。

解密是加密的逆过程。加密变换必须对明文和密钥敏感，同时能够最大程度地改变明文统计特性。为此，依据 Shannon 提出的密码系统设计的两个基本原理——混乱和扩散，把算法分为两个模块：像素替换和位置置换。前者起到混乱作用；后者起扩散作用。两部分的设计都符合模块化设计思想，实现方便。基于该理论，提出如下解决方案：

第一个模块是像素替换模块。考虑一幅大小为  $M \times N$  具有  $L$  级灰度的图像，设  $I(i, j)$  为  $(i, j)$  坐标处图像的像素值，其中  $1 \leq i \leq M, 1 \leq j \leq N, I'(i, j)$  为  $(i, j)$  坐标处替换操作后图像的像素值，即要求设计混沌映射  $f$ ，使得：

$$f : I(i, j) \rightarrow I'(i, j) \tag{10-6}$$

为了使替换加密后的像素值  $I'(i, j)$  具有不可预测性，采用指数映射和正弦平方映射来控制替换操作。

第二是位置置换模块。是将像素的次序按混沌特性打乱，使  $(i, j) \rightarrow (i', j')$ ， $1 \leq i \leq M, 1 \leq j \leq N$ ，使图像变得“面目全非”，但像素的总个数不变，这样能有效地抵御统计和预测分析。

设计好两个模块，就可以实现前面提到的加密/解密算法设计。加密算法先经过像素替换模块，再经过位置置换模块，实现对明文图像（原始图像）的加密。其中像素替换模块主要起混乱作用，而位置置换模块主要起扩散作用，经过这样的变换，使得输出密文对输入明文和密钥充分敏感。解密时与加密时顺序相反，即先经过逆位置置换模块，再经过逆像素替换模块，从而实现图像密文的解密。算法的加密解密框图如图 10-14 所示。

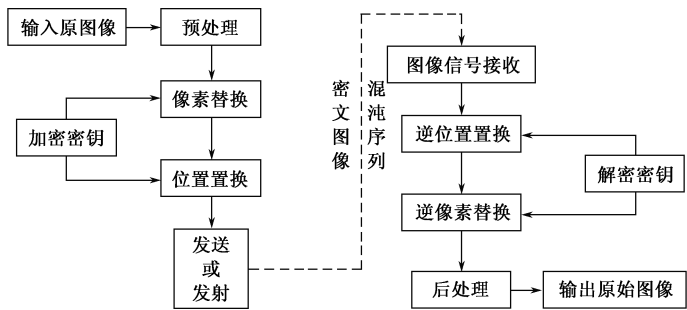


图 10-14 加密/解密算法框图

由图 10-14 可以看出，加密和解密的结构是不同的，加密时先进行预处理，包括预填充和图像分块，而解密最后要再进行处理，包括图像块重组和截取。整个算法符合模块化设计，实现方便，而且两个模块都是应用混沌系统来设计。因此，每个模块都具有内在的随机性，利于密码学应用。

图 10-15 给出了一组原始图像，经过像素替换和位置置换后的密文图像和解密后图像。

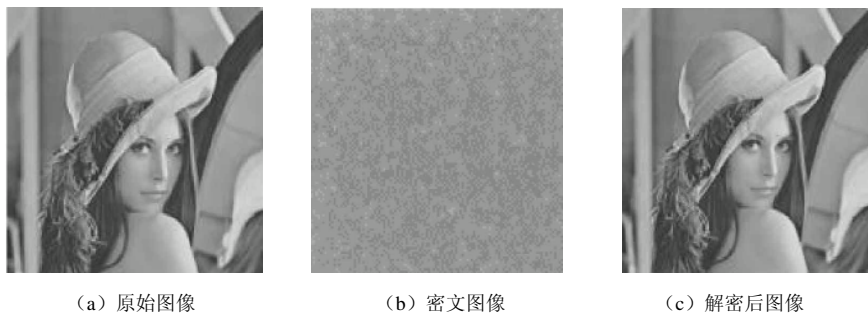


图 10-15 图像加密处理示意

另外一种常见的图像加密技术，是基于时空混沌序列算法，是将混沌耦合映像格子看做图像基本单元，提出使用时空混沌进行图像加密和解密方案。基本思路是发送端基于图像信息产生时空混沌序列，并用它加密压缩图像；在接收端产生同样的时空混沌序列，恢复出原始图像。

除了以上两种图像加密方法，最常用的还有杜芬型混沌序列图像数据掩盖加密，这里就不具体介绍了。

10.3.3 数据通信网络保密技术

随着电子技术、通信技术和计算机技术的快速发展，通信系统已由传统的模拟体制过渡到目前的全数字（数据）通信体制，数据通信 OSI 分层的出现，使数据通信网也进入到网络保密阶段。

在 OSI 安全体系结构中，根据不同的用户需求，通常在 OSI 的不同层上进行加密。比如说，为了保护通信的正文，在表示层或应用层对数据加密即可；若要对网络口令、控制信息等进行保护，并防止业务流分析，则必须在物理层进行加密才能达到目的；若用户要求只对源节点到目的节点之间传输的信息进行保护，则在传输层增加加密措施即可。下面我们就介绍常见的几种网络保密技术。

1. 端对端加密方式

网络用户终端对终端加密是对一对用户之间的信息提供保护。它要求各对用户采用相同的密钥。对于传送通路上的各中间节点，信息是保密的，如图 10-16 所示。

计算机互联网中，终端保密通信是端到端通信保密，是在两个主机应用之间端到端链接的数据传输保密，而不是主机之间通信底层的点对点保密通信，端到端通信保密只能在传输层或传输层以上的高层协议中完成，而不能在传输层以下协议层实现。因为网络层协议控制信息对网络层数据单元通过的实际节点情况是不可知的，如果对网络层协议控制信息的加密，



网络中间节点就无法实现数据报正常的路由选择。如果选择在传输层完成加密，则除运输层控制信息外，传输层以上其他各层的控制信息都暴露了，容易受到攻击。如选择在应用层完成加密，可以实现应用到应用的安全保护；两端用户可以根据自己的特殊需求，独立选择加密算法和灵活更换密钥，和其他用户无关，从而具有较高的安全性，但缺乏通用性。

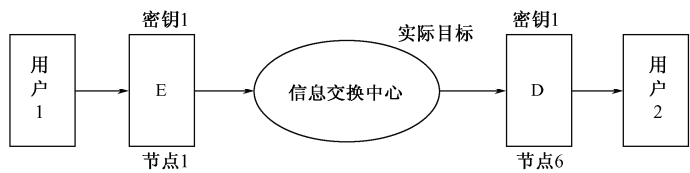


图 10-16 端对端通信加密示意图

由于网络中间节点要进行路由选择，端到端通信保密中网络层及以下各层的控制信息如源地址、目的地地址、路由信息等都不能加密。这给攻击者进行攻击提供了方便。

2. 链路加密和群路加密方式

链路加密是对网中两个相邻节点之间传输的数据进行加密保护，即节点对节点加密，这不同于终端加密中的点对点通信加密。在链路加密情况下，用户端和用户端之间可能有多个节点。

对于链路加密，除了数据单元外，链路层以上的所有控制信息（源地址、目的地地址和连接信息等）都被加密，所以在下一个节点处必须解密才能完成正确的路由选择，如图 10-17 所示。

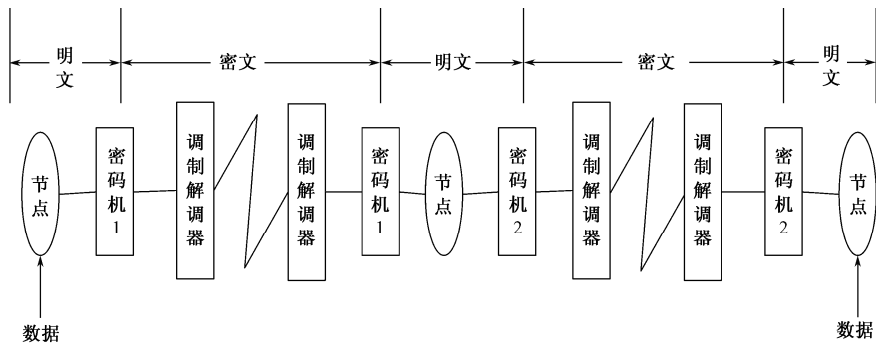


图 10-17 逐链加密

通信网的链路加密是在节点之间每段独立链路的加密处理，所以每段链路的加密密钥可以是不同的，甚至可以是不同的加密体制。而一段链路的破译不会影响其他链路的安全。在受保护数据所选定的路由上，任一对节点和相应的调制解调器之间安装有相应的密码机，并配置相应的密钥，不同节点对之间的密码机和密钥不一定相同。由于信息在各节点内重新完成加解密，所以节点有信息的明文存在，这要求节点必须是安全的，一旦中间主要节点的失控将暴露所有信息。对采用自动路由的网络，攻击者掌握一个节点就可以设法修改路由控制，使感兴趣的信息经过该节点。链路加密虽然能防止搭线窃听，但不能防止在消息交换过程中由于错误路由造成的泄密。

群路保密是对综合传输体系中多链路复用后的高速（宽带）信号进行加密，广泛使用在通信网干线上。如传输体系中多话音链路复用后加密等，特别使用在卫星、微波、无线等信号比较容易截收的干线信道上。除使用在通信干线外，在重要的办公楼群与交换局之间，用户信号也可复用后集中加密，完成主要政府办公楼与交换局用户线的群路集中加密和解密。

链路加密条件下，由网络提供密码功能，故对用户来说是透明的。在端对端加密的条件下，如果加密功能由网络自动提供，则对用户来说也是透明的；如果加密功能由用户自己选定，则对用户来说就不是透明的。端对端加密方式只在需要加密保护信息的用户之间备有密码设备，从而大大减少了整个网中使用密码设备的数量。

3. 两种加密方式的比较

采用链路加密方式，从起点到终点，要经过许多中间节点，在每个节点均要暴露明文（节点加密方法除外），如果链路上的某一节点安全防护比较薄弱，那么按照木桶原理（木桶水量是由最低一块木板决定），虽然采取了加密措施，但整个链路的安全只相当于最薄弱的节点处的安全状况。采用端端加密方式，只是发送方加密报文，接收方解密报文，中间节点不必加解密，也就不需要密码装置。此外，加密可采用软件实现，使用起来很方便。在端端加密方式下，每对用户之间都存在一条虚拟的保密信道，每对用户应共享密钥（传统密码保密体制，非公钥体制下），所需的密钥总数等于用户对数目。对于  $n$  个用户，若两两通信，共需密钥  $n \times (n-1)/2$  种，每个用户需  $(n-1)$  种。这个数目将随网上通信用户的增加而增加。为安全起见，每隔一段时间还要更换密钥，有时甚至只能使用一次密钥，密钥的用量很大。

链路加密，每条物理链路上，不管用户多少，可使用一种密钥。在极限情况下，每个节点都与另外一个单独的节点相连，密钥的数目也只是  $n \times (n-1)$  种。这里  $n$  是节点数而非用户数，一个节点一般有多个用户。从身份认证的角度看，链路加密只能认证节点，而不是用户。使用节点 A 密钥的报文仅仅保证它来自节点 A。报文可能来自 A 的任何用户，也可能来自另一个路过节点 A 的用户。因此链路加密不能提供用户鉴别。端端加密对用户是可见的，可以看到加密后的结果，起点、终点很明确，可以进行用户认证。网络中传送的消息由报头（含目的地、作业号、报文源、起止指示符、报文类别、格式等业务数据）和报文（用户之间交换的数据）组成。在链路加密下，报文和报头可同时进行加密，这有利于对抗业务流量分析。表 10-2 对链路加密和端端加密进行了比较。

表 10-2 逐链加密和端端加密的比较

|      | 优 点   | 缺 点   |
|------|---|---|
| 逐链加密 | 1. 所有信息都被加密，包括报头和路由器；<br>2. 单个密钥的泄密不会使整个网络泄密；<br>3. 加密对用户是透明的；<br>4. 所需密钥量比端一端加密少                   | 1. 报文以明文方式通过每个节点；<br>2. 由于所有网络节点都要收密钥；<br>3. 每条保密通信链路都需配两台密码设备；<br>4. 仅仅能对通信的节点进行验证           |
| 端端加密 | 1. 比较灵活。加密可置于用户控制之下，并且不必对所有信息都加密；<br>2. 数据在网上从源点到目的点都受到保护；<br>3. 对网络节点来说，加密是透明的；<br>4. 可对通信方的身份予以确认 | 1. 要求每个系统都双月刊进行同样的加密；<br>2. 某些信息（如报头和路由器信息）可能会以明文的方式传送，将难以防止信息流分析攻击；<br>3. 要求比较复杂的密钥分配和密钥管理技术 |

#### 4. 混合加密方式

在端一端加密下，只能对报文加密，报头则以明文形式传送，容易遭受业务流量分析。为了保护报头中的敏感消息，可以采用图 10-18 所示的端端和链路混合加密方式。在此方式下，报文将被两次加密，第一次用户 1 用密钥 1 加密，而第二次是用密钥 2 加密，而在进入交换前把第二次加密的解密下来，这第二次加密和解密叫做链路方式加密。同样图中用密钥 3 和密钥 4 的加密和解密也是链路方式加密，而报文的报头则只由链路方式进行加密，报文的其他部分使用端端方式加密。这样，在两端用户间的通信链路上，中间各节点因为采用链路加密即可获得路由信息，而报文其他部分因为采用端端加密而不会出现明文形式。

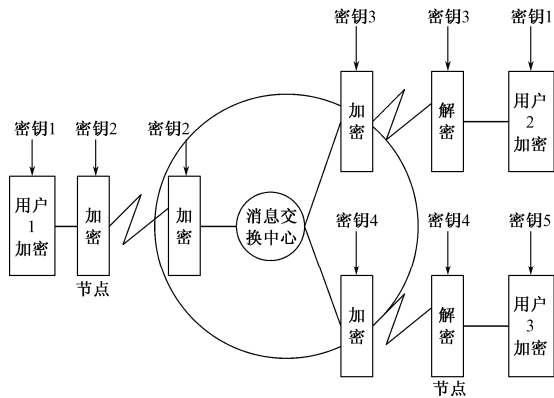


图 10-18 混合加密方式

在明文和密文的混传网络中，可在报头的某个特定位上指示报文是否被加密，也可按线路协议由专用控制信息实现自动起止加密操作。从成本、灵活性和安全性来看，一般端对端加密方式较有吸引力。对某些远程处理机构，链路加密可能更为合适，当链路中节点数很少时，逐链加密操作对现有程序是透明的，无须操作员干预。目前大多数链路加密设备是以线路的工作速度工作，因而不会引起传输性能显著下降。另外，有些远端设备的设计或管理方法不支持端一端加密方式。端一端加密的目的是对从数据的源节点到目的节点的整个通路上所传的数据进行保护，而链路加密的目的可对全部通路或链路中潜伏截收危险的一段通路进行保护。网中所选用的数据加密设备要与数据终端设备和数据电路端接设备的接口一致，并且要遵守国家和国际标准规定。

#### 思考与练习

- 10-1 试画图并说明密码体制的一般模型。
- 10-2 何谓理论（完全）保密和实际保密？
- 10-3 比较非对称密码体制与对称密码体制的优缺点。
- 10-4 试说明穷举破译法和分析破译法的主要区别。
- 10-5 何谓图像保密通信？试画出图像保密通信系统的组成框图。
- 10-6 简述几种模拟语音信号加密方法，与数字语音有何区别。
- 10-7 简述网络终端加密与链路加密的区别。

# 第 11 章 无线数据通信协议

IEEE (Institute of Electrical and Electronics Engineers, 电气电子工程师学会) 802 标准组根据覆盖范围, 将无线数据网络划分为通信距离小于 10 m 的无线个域网 (Wireless Personal Area Network, WPAN)、通信距离在 100 m 以内的无线局域网 (Wireless Local Area Network, WLAN)、可涵盖城市或郊区等较大地理区域的无线城域网 (Wireless Metropolitan Area Network, WMAN)、涵盖面积最大的无线广域网 (Wireless Wide Area Network, WWAN)。

在无线数据通信网中, 信道特性的非对称性、长时延、数据包丢失及比特差错的严重性等, 使其有别于有线通信环境, 其协议或规程不能直接照搬, 因此编写本章内容。

## 11.1 数据通信协议的概念

### 11.1.1 协议的概念及功能

数据通信协议是网络内使用的“语言”, 用来协调网络的运行, 以达到互通、互控和互换的目的, 通信的双方要共同遵守这些约定。在现代通信网中, 一般用户通过应用软件使用数据信息, 应用软件使用网络进行通信时并不直接同网络硬件打交道, 而是同给定的协议规则打交道, 通信网络协议是通信网络中不可缺少的重要组成部分。

#### 1. 通信网络协议组成要素

一个通信网络协议主要由以下三要素组成:

- (1) 语法——规定通信的双方以什么方式交流数据信息, 即确定数据与控制信息的结构或格式, 其内容包括信息传送的格式、接口标准、差错控制的方式等。
- (2) 语义——规定通信的双方要交流哪些数据信息, 即确定需要发出何种控制信息, 完成何种动作, 以及返回什么应答等。
- (3) 定时关系——规定事件执行的顺序, 即确定通信过程中的状态变化, 对发送的信息进行编号, 以免重复接收或丢失。

#### 2. 通信网络协议主要功能

数据通信主要是人与机、机与机之间的通信, 因此通信协议应规范得十分详尽才能保证通信的正常进行。通信协议是一个复杂而庞大的通信规则的集合, 它应完成以下 8 种主要功能:

- (1) 信号的传送与接收。应规定的内容包括信息传送的格式、接口标准及启动控制、超时控制等功能。
- (2) 差错控制。应使终端输出的数据具有一定的差错控制功能, 目的终端根据收到的数据可进行相应的检错或纠错操作。
- (3) 顺序控制。对发送的信息进行编号, 以免重复接收或丢失。
- (4) 透明性。透明性是指对用户端所使用的代码无任何约束性的限制, 也即应采取必要的措施, 保证所传送的数据信息为随机的比特序列。

(5) 链路控制与管理。应控制信息的传输方向、建立和结束链路的逻辑连接, 显示数据终端设备的工作状态等。

(6) 流量控制。为了避免链路阻塞, 应能调节数据链路上的信息流量, 能够决定暂停或继续接收信息。

(7) 路径选择。确定信息报文通过多个节点和链路到达目的节点的传输路径和最优的路径选择策略。

(8) 对话控制。包括信息的处理、信息安全和保密、应用服务等内容。

### 3. 功能与协议分层结构的好处

数据通信协议是一个复杂而庞大的通信规则的集合, 采用分层结构, 它把实现通信的网络在功能上视为若干相邻的层组成, 每一层完成其特有的功能。每一层都建立在较低层的基础上, 利用较低层的服务, 同时为较高一层提供服务。网络功能上的分层必然导致协议的分层即把复杂的协议分解为一些简单的协议, 再组合成总的协议。

分层的好处有:

(1) 各层相互独立, 需要知道的仅是层间接口所提供的服务。一层是一种类型的功能的集合, 它由许多功能块组成, 每一个功能块执行协议规定的一部分功能;

接口是指穿越相邻层之间界面进行数据传送的一组规则, 它可以是硬件接口, 也可以是软件接口, 如数据格式的转换等; 服务是指某一层及其以下各层通过接口提供给上层的一种能力。一层包含一系列的服务, 每个服务则是通过某一个或某几个协议来实现, 而相邻层之间通过接口通信。

(2) 灵活性好。当任何一层发生变化时, 只要接口关系保持不变, 则上下相连的层均不受影响, 且层内提供的服务可修改; 当提供的服务不再需要时, 可将该层取消。

(3) 结构上独立, 各层可采用最合适的技术来实现。

(4) 方便实现和维护。

(5) 促使标准化工作。

## 11.1.2 开放系统互连参考模型

### 1. 基本概念

国际标准化组织 (ISO) 于 1977 年提出了开放系统互连 (OSI) 参考模型, 它是一个异种计算机连接标准的框架结构。OSI 为连接分布式应用处理的“开放”系统提供了基础。

所谓“开放”是强调对 OSI 标准的遵从, 即只要遵循 OSI 标准的系统, 就可以互相通信, 即具有开放性。

所谓“系统”是指实际通信系统中与 OSI 有关的各部分, 是一个或多个计算机、相应的软件、外围设备、终端、终端操作员、物理过程及信息传送手段等的集合。

开放系统是指在与其它系统通信时, 遵守 OSI 标准要求的系统。

OSI 涉及的是为完成一个公共 (分布的) 任务而相互配合的系统能力及开放式系统之间的信息交换, 但它不涉及系统的内部功能和与系统互连无关的其他方面, 也就是说系统的外部特性必须符合 OSI 的网络体系结构, 而其内部功能不受此限制。采用分层结构的开放系统

互连大大降低了系统间信息传递的复杂性。

应当理解 OSI 参考模型仅仅是一个概念性和功能性结构，它并不涉及任何特定系统互连的具体实现、技术或方法。

OSI 参考模型如图 11-1 所示，共有 7 个功能层，自下而上分别是：物理层、数据链路层、网络层、运输层、会话层、表示层、应用层。分别用各层英文首字母缩写 PH、DL、N、T、S、P、和 A 表示。

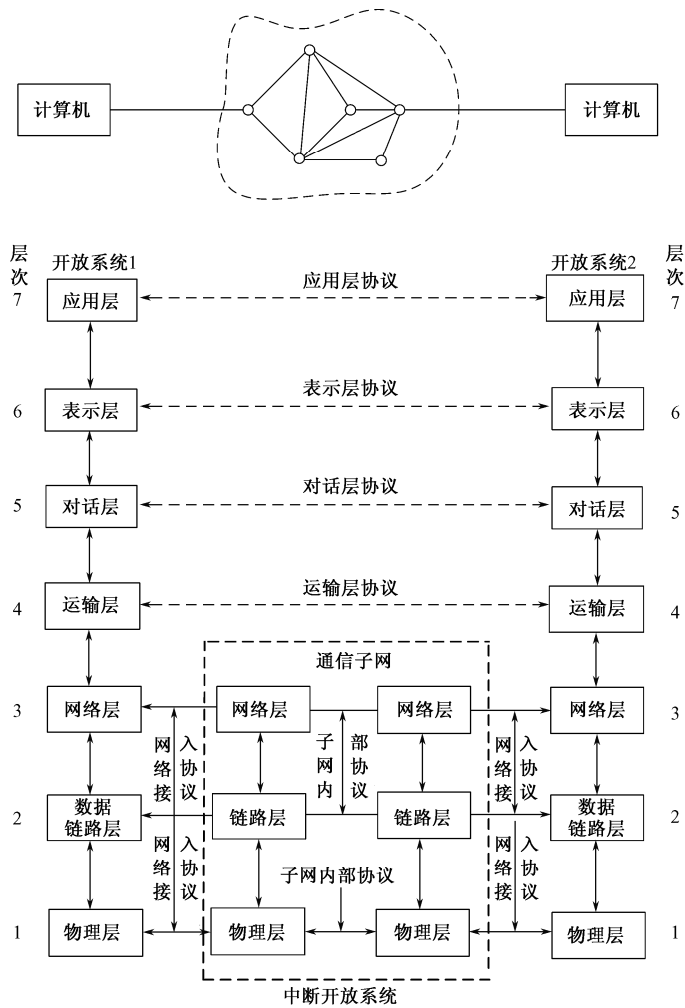


图 11-1 开放系统互连（OSI）参考模型

OSI 模型所描述的范围包括端开放系统与中继开放系统。中继开放系统即指图 11-1 中通信子网部分，连接点的物理传输介质不在 OSI 范围内。

## 2. 各层功能概述

### 1) 物理层

物理层并不是无线或有线物理媒介本身，它是开放系统利用物理媒介实现物理连接的功能。

能描述和执行连接的规程。

物理层提供用于建立、保持和断开物理连接的电气的、机械的、功能的和规程的特性。保证比特流的正确传输，提供有关同步和全双工比特流在物理媒介上的传输手段。

对于无线数据系统来说，物理层是为高层信息提供的无线信道，即无线物理媒介上传输信息所需要的全部功能，如极化方式、调制与解调方式、频率分配、信道划分、传输定时、加密、功率设定和小区划分等。（其主要内容在本书前面已经有所介绍）

典型的 GSM 系统中的广播信道（BCH）、随机接入信道（RACH），空间数据系统咨询委员会（CCSDS）邻近空间链路物理层特性 Prox-1 中，详细规定了物理层中的极化方式、调制与解调方式、频率分配、信道划分、传输定时、加密、功率设定和小区划分等功能。

3G 网络的物理层关键技术有 OFDMA/MIMO/AAS/AMC/LDPC/HARQ 等。

## 2) 数据链路层

本层的主要功能是将物理层提供的物理传输信道处理成可靠的信道，即将数据组成适合于正确传输的帧形式的数据块。

如无线数据链路的建立与维护、确认模式的帧传送与接收、信道接入控制、帧校验、预留时隙管理、广播信息管理。

常用的数据链路层协议有基本型传输控制规程和高级数据链路控制规程（HDLC）等。

## 3) 网络层

计算机（或数据终端）网络是由一系列用户终端、主机和具有信息处理和交换功能的节点（即交换机）、以及连接它们的传输线路组成。所以通常把计算机（或数据终端）网分成用户子网（资源子网）和通信子网（即数据通信网）两部分。用户子网向用户提供访问网络的机会，通信子网则提供网络通信的能力。

网络层提供系统之间网络的连接，它负责将两个终端系统，经过网络节点用数据链路连接起来，完成一条通信通路。由于传输设施是多种类型的，有的用卫星租用线路、光缆租用线路、有的用交换线；有点到点、有虚电路及分组网等。网络层只用于控制通信子网的运行，它的存在使网络布局和业务上的差异不会被其高层所察觉，换言之，在它的高层来看是透明的。上一层不需要了解网络中数据传输和交换的细节。

网络层规定了网络连接的建立、拆除和通信管理的协议，通信管理包括数据交换、路由选择、通信流量控制等。

网络层在各节点间建立可靠的数据链路基础上，实现源节点到目的节点间建立一条传输通道。具体的功能有控制分组传送系统操作的路由选择、拥塞控制、网络互连等；根据传输层的要求来选择服务质量；向传输层报告未恢复的差错等。

在无线数据通信中，网络层主要功能是管理链路连接、控制呼叫过程、支持附加业务和短信息业务，以及进行移动管理和无线资源管理等。

## 4) 运输层

运输层也称传输层，是开放系统之间建立、维持和拆除传送控制层（注意：链路层中，仅保证节点间传输控制），实现用户的端到端的或进程之间数据的透明传送，保证端到端数据无差错、无丢失、无重复、无次序（分组）颠倒等的顺序控制、流量控制，使会话层实体不

需要关心数据传送的细节，同时，还用于弥补各种通信子网的质量差异，对经过低 3 层仍然存在的传输差错进行差错控制、差错恢复。另外，该层给予用户一些监督选择，以便从网络获得某种等级的通信服务质量。总体来说，它是面向通信子网的低层和面向数据处理高层的接口。

### 5) 会话层

为了两个进程之间的协作，必须在两个进程之间建立一个逻辑上的联系，这种逻辑上的连接称之为会话。会话层作为用户进入传输层的接口，负责进程间建立会话和终止会话，并且控制会话期间的对话，提供诸如会话建立时会话双方资格的核实和验证，由哪一方支付通信费用，及时对对话方向的交替管理、故障点定位和恢复等各种服务。它提供一种经过组织的方法在用户之间交换数据。“会话”的方式有双方同时进行、交替进行或单向进行。

### 6) 表示层

表示层提供数据的表示方法，其主要功能是从应用层接收数据类型（字符、整数等），并将其转换成通信对象的对等层共同理解的句法，包括字符代码（如 ASCII）、数据格式、控制信息格式、加密等。即功能是处理两个应用实体之间进行数据交换的语法变换及数据结构不一致及数据不同编码的转换，并提出数据压缩和数据加密功能来实现数据交换的可能性和可靠性。

### 7) 应用层

直接对用户提供服务，它为用户进入系统互连提供一个窗口（工具），实现终端用户应用进程之间的信息交换。与表示层不同，它关心的是数据的语义。应用层包含管理功能，提供公用的应用程序，如作业管理、金融数据交换、电子信箱、数据库管理、事务处理、文件传送和控制和网络管理等。由于网络应用的要求很多，所以应用层最复杂，所制定的应用协议也最多。

在 7 层协议中，下面 3 层（物理层、数据链路层和网络层）称为通信层或无线空中接口层，上面 4 层称为用户层。一般来说，网络的低层协议决定了一个网络系统的传输特性，如所采用的传输介质、拓扑结构及介质访问控制方法等，通常由硬件来实现。高层协议提供了与网络硬件结构无关的，更加完善的网络服务和应用环境，通常是由网络操作系统实现的。

实际系统通信过程也并非都是经过 7 层，大多只有低 4 层，或者低 3 层。另外，在具有功能层次相同的两个不同系统上进行通信是在对等层进行的，这种通信称为虚通信。之所以成为虚通信，是因为对等层与对等层之间不能直接进行物理通信，只有通过其下层直至最下面的物理层来传输数据。

## 11.2 数据链路传输控制规程

### 11.2.1 数据链路控制规程的功能与分类

#### 1. 数据链路的概念

国际标准化组织（ISO）给数据链路的确切定义为：按照信息的特定方式进行操作的两



个或两个以上的终端装置与互连线路的一种组合体。所谓特定方式是指信息速率和编码均相同。

数据通信就要在数据电路已建立的基础上，通过两端的控制装置使发送方和接收方之间交换握手信号，双方确认后可开始传输数据。

数据链路的结构有点对点和点对多点两种，从链路逻辑功能的角度看，在数据电路上发送和接收数据的数据终端设备统称为站。

在点对点链路中，有主站、从站和组合站之分，主站是发送信息和命令的站、从站是接收信息和命令并发出认可信息或响应的站，组合站是兼有主站和从站功能的站，如图 11-2 (a) 所示。

在点对多点链路中，有控制站、辅助站、从站、主站和中性站之分。负责组织链路上的数据流，并处理链路上不可恢复差错的站称为控制站，其余各站称辅助站。如果控制站发送信息，辅助站接收信息，则此时控制站称主站，接收信息的辅助站称从站，其余辅助站称中性站，如图 11-2 (b) 所示。

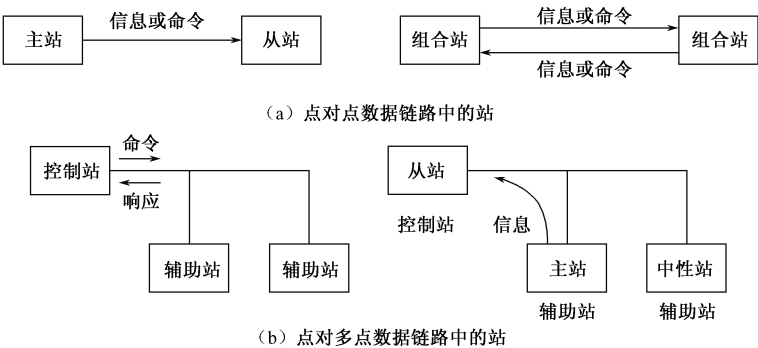


图 11-2 数据链路的结构

数据链路传输数据信息有 3 中不同的操作方式：信息只按一个方向传输的单项型(单工)；信息从一个方向传播，然后再从相反方向传输的双向交替型(半双工)；信息可在两个方向上同时传输的双向同时型(双工)。

为了在 DTE 与网络之间或 DTE 与 DTE 之间有效、可靠地传输数据信息，必须在数据链路这一层次上采取必要的控制手段对数据信息的传输进行控制。传输控制是通过数据链路层协议来完成的，习惯上把数据链路层协议称为数据链路传输控制规程。

为了充分理解数据链路控制规程的功能，我们先来分析一下一次数据通信的过程。与电话通信过程类似，一次完整的数据通信过程包括以下 5 个阶段：

### 1) 建立物理连接

物理连接就是物理层若干数据电路的互连。数据电路有交换型和专用型线路两种。专用型线路为通信双方提供永久性固定的连接，因此无须“建立物理连接”。交换型数据电路必须按交换网要求进行呼叫连接。在电路交换网中，通常要求发送方完成自动存取号码、拨号、发送和应答数据单音、控制呼叫应答器与调制解调器之间的线路转换等操作。该阶段与电话通信的呼叫建立阶段类似。如电话网的 V.25 和数据网的 X.21 呼叫应答规程。

## 2) 建立数据链路

该阶段类似于在电话通信中建立起物理连接之后，相互证实的过程。如主叫方询问被叫方“你是XXX？”，如果是要找的对象，双方就进入通话阶段。而在数据通信的这个阶段，为了能可靠而有效地传输数据信息，收发双方也要交换一些控制信息，包括呼叫对方、确认对方是否所要通信的对象；确定接收和发送状态；哪一方为发送状态，哪一方为接收状态；指定双方的输入/输出设备。

由于数据链路的结构不同，传输控制的方式也不同，因此必须采用不同的方式来建立数据链路。建立数据链路的方式有两种：争用方式和探询/选择方式。争用方式是争做主站建立链路的一种方式，探询/选择方式是在有控制站系统中，为防止争夺做主站，控制站按一定顺序一次一个地引导辅助站变成主站（探询过程），主站用选择序列引导一个或多个从站接收报文（选择过程）的方式。一般而言，对于两点直通式的链路结构常采用争用方式，有时也采用探询/选择方式，但多点式的链路结构通常只用探询/选择方式。

## 3) 数据传送

该阶段类似于电话通信的通话阶段。电话通信中双方要进行通话，首先必须使用相同的语言，否则要借助翻译；其次双方要相互配合，说和听要有一定的顺序，不能同时讲和连续讲而不管对方是否听懂；另外如果没有听清或漏听对方的话，要求对方重讲，直到听清为止。数据的传送也类似，主站按规定的格式组织数据信息，并按规定的顺序沿所建立的数据链路向从站发送，同时完成差错控制、流量控制等功能，使数据信息有效、可靠、透明地传输。

## 4) 释放数据链路

对电话通信，当确认双方均无信息传送之后进入通信结束阶段。在数据信息传送结束后，主站发送通过规定的结束字符（结束传输的命令）来拆除数据链路，各站返回中性状态，数据链路被拆除。需注意的是拆除数据链路并不是拆除物理连接，该阶段后可以又一次进入第二阶段，建立新的数据链路（即一个数据通信系统可以建立一个或多个数据链路），这就类似于主叫方同一个人讲完话后，可能还想同另一个人讲话，则进入第二阶段。

## 5) 拆除物理连接

电话通信中，通话结束后，任何一方挂机，交换网络就拆除物理连接。数据通信中，当数据链路的物理连接是交换型电路时，数据传送结束后，只要任何一方发出拆线信号，便可拆除通信线路，双方数据终端恢复到初始状态。

以上5个阶段，第2到第4阶段属于数据链路控制规程的范围，而第1和第5阶段是在公用交换网上完成的操作。

## 2. 数据链路控制规程的功能

链路层协议在数据通信中称为数据链路控制规程，它应具备如下功能：

### 1) 帧控制

在数据链路中的数据是以一定长度和一定格式的信息块形式传送的，这种信息块称为

帧。不同的应用，帧的长度和格式可以不同。控制规程对帧的类型和结构进行了规定，指出每帧由一些子段和标志组成，标志用于指明帧的开始和结束，字段则根据不同用途分为地址字段、控制信息字段和校验字段等。

在传送方，帧控制具有把从上层来的数据信息分成若干组，并在这些组中加入标志和字段组成一帧功能，在接收方，帧控制具有把接收到的帧去掉标志和字段，还原原始数据信息后送到上层的功能。

### 2) 透明传送

控制规程中应采取必要的措施，保证所传送的数据信息不能出现与标识和字段相同的组合，如果出现，要求采取打乱措施，使用户传输信息不受限制。

### 3) 差错控制

由于物理电路中存在各种各样的干扰，数据信息在传送过程中不可避免地要产生差错。控制规程应能控制电路中产生的差错。常用的差错控制方式有检错重发和前向纠错两种。为防止帧的重收和漏收，必须采用帧编号发送，接收时按编号认可。当检测出无法恢复的差错时，应通知网络层实体作相应的处理。

### 4) 流量控制

为了避免链路阻塞，控制规程应能调节数据链路上的信息流量，能够决定暂停、停止或继续接收信息。常用流量控制的方法有发送等待方法、面向帧控制方向、滑动窗口控制等多种方法。

### 5) 链路管理

控制信息传输的方向，建立和结束链路的逻辑连接，显示数据终端设备的工作状态等。

### 6) 异常状态的恢复

当链路发生异常情况时，能够自动地恢复到正常状态。

## 3. 数据链路控制规程的分类

根据帧格式，数据链路控制规程有两种：面向字符型和面向比特型。

典型的面向字符的控制规程有 IBM 公司的二进制同步通信规程 (BISUNC 或 BSC)、ISO 的基本型控制规程 (ISO-1745 标准)、美国国家标准协会 (ANSI) 规定的 ASSI X3.28 通信控制规程、欧洲计算机协会 (ECMA) 的 ECMA-16 基本型控制规程和我国数据通信基本型控制规程 (GB3453—82) 等。由于这种通信规程与特定的字符编码集关系过于密切，兼容性较差，并且在实现上也比较复杂，故现代的数据通信系统已很少使用。

面向比特型的控制规程有 IBM 公司的同步数据链路控制规程 (SDLC)、ANSI 的先进数据通信控制规程 (ADCCP)、ISO 的高级数据链路控制规程 (HDLC) 和 CCITT 的 X.25 建议中的链路地 HDLC 的变种等。这些规程是目前通信网中最常用的通信控制规程。

11.2.2 面向字符型的传输控制规程

1. 规程基本特征

面向字符型控制规程最基本的特征是：

- 字符编码采用 CCITT 建议的国际 5 号编码表；
- 以字符为最小控制单位，它规定了 10 个控制字符用于传输控制，如表 11-1 所示；
- 通信方式为双向交替型（半双工）；
- 可采用起止式的异步传输方式和同步传输方式；
- 检错采用行列监督码；
- 差错控制方式采用检错重发（ARQ）的纠错方式。

表 11-1 控制字符

| 类 别    | 名 称    | 字符（英文名称）                        | 功 能             |
|--------|--------|---------------------------------|-----------------|
| 格式字符   | 标题开始   | SOH (Start of Head)             | 表示信息报（电）文标题的开始  |
|        | 正文开始   | STX (Start of Text)             | 表示信息报（电）文正文开始   |
|        | 正文结束   | ETX (End of Text)               | 表示信息报（电）文正文结束   |
|        | 码组传输结束 | ETB (End of Transmission Block) | 正文码组结束          |
| 基本控制字符 | 询问     | ENQ (Enquiry)                   | 询问对方要求回答        |
|        | 传输结束   | END (End of Transmission)       | 表示数据传输结束        |
|        | 确认     | ACK (Acknowledge)               | 对询问的肯定回答        |
|        | 否定回答   | NAK (Negative Acknowledge)      | 对询问的否定回答        |
|        | 同步     | SYN (Synchronous Idle)          | 用于建立同步          |
|        | 数据链路转义 | DLE (Data Link Escape)          | 用来与后继字符一起组成控制功能 |

2. 报文格式

报文格式也叫文电格式，就是利用上述编码字符进行数据传输时所规定的排列形式。报文分为两类：信息报文（信息文电）和监控报文（监控序列）。

1) 信息报文（信息文电）格式

信息报文的 4 种格式如图 11-3 所示。信息报文包括正文和标题（报头）。正文是要传输的字符信息。标题是与报文正文的传送和处理相关的一组辅助字符信息，它包括发信地址、收信地址、优先权、保密措施、信息报文名称、报文级别、编号、传输路径等。

标题以字符 SOH 开始，正文以字符 STX 开始，以字符 ETX 结束。正文紧接在标题后面。

正文的长度没有限制，为了便于差错控制，可以把正文分成若干个码组（Block），码组的长度取决于数据电路的传输质量。一个码组以字符 STX 开始，以字符 ETB 结束，对于正文的最后一个码组以字符 ETX 结束。在每个码组或正文的后面为码组校验字符（BCC）。

在上述格式中，信息的传输是不透明的，它不允许在正文或标题中出现与控制字符相同的序列，这样大大限制了它的应用。为了能透明传输，必须在控制字符 SOH、STX 等前面加上转义字符 DLE 来实现。如用 DLE SOH 表示标题开始，DLE STX 表示正文开始等。这样当

信息出现与控制字符相同的序列时，因为前面没有转义字符 **DLE** 而不会被误解。但是，如果在信息中出现与 **DLE** 相同的字符序列时，会出现特殊的问题，解决的办法是在该序列前面再加上一个 **DLE** 后发送，在接收端自动删除额外附加的 **DLE**。

|     |    |     |        |     |     |
|-----|----|-----|--------|-----|-----|
| SOH | 标题 | STX | 信息报文正文 | ETX | 校验码 |
|-----|----|-----|--------|-----|-----|

(a) 信息报文的基本格式

|     |         |     |     |
|-----|---------|-----|-----|
| STX | 信息报文正文组 | ETX | 校验码 |
|-----|---------|-----|-----|

(b) 无标题的信息报文格式

|     |    |     |          |     |     |
|-----|----|-----|----------|-----|-----|
| SOH | 标题 | STX | 信息报文正文组1 | ETB | 校验码 |
|-----|----|-----|----------|-----|-----|

|     |          |     |     |
|-----|----------|-----|-----|
| STX | 信息报文正文组2 | ETB | 校验码 |
|-----|----------|-----|-----|

⋮

|     |             |     |     |
|-----|-------------|-----|-----|
| STX | 信息报文正文组 $n$ | ETX | 校验码 |
|-----|-------------|-----|-----|

(c) 将信息报文分成 $n$ 个信息报文的格式

|     |      |     |     |
|-----|------|-----|-----|
| SOH | 标题组1 | ETB | 校验码 |
|-----|------|-----|-----|

|     |      |     |     |
|-----|------|-----|-----|
| SOH | 标题组2 | ETB | 校验码 |
|-----|------|-----|-----|

⋮

|     |         |     |        |     |     |
|-----|---------|-----|--------|-----|-----|
| SOH | 标题组 $m$ | STX | 信息报文正文 | ETX | 校验码 |
|-----|---------|-----|--------|-----|-----|

(d) 将信息报文的标题分成 $m$ 个组的格式

图 11-3 信息报文的 4 种格式

2) 监控报文（监控序列）格式

监控报文一般由单个传输控制字符或由若干图形字符引导的单个传输控制字符组成。引导字符统称为前缀，前缀的长度不超过 15 个字符，它包含标识信息、地址信息、状态信息以及其他通信控制所需的信息。监控报文按传输方向可分为两种：正向监控报文（正向监控序列）和反向监控报文（反向监控序列）。

正向监控报文与信息报文方向一致，是主站向从站发送的控制序列。主要用于通信双方的呼叫应答，以确保信息报文的可靠传输。正向监控报文的形式如表 11-2 所示。在同步传输中把两个或多个 **SYN** 序列放在前面，以建立和维持收发两站的同步。

表 11-2 正向监控报文

| 名 称        |          | 控 制 报 文     |
|------------|----------|-------------|
| 探询报文       |          | 探询地址 ENQ    |
| 选择报文       | 选择站      | 选择地址 ENQ    |
|            | 选择地址 ENQ | （前缀）ENQ     |
|            | 建立数据链路   | （前缀）ENQ     |
| 询问报文（催促应答） |          | （前缀）ENQ     |
| 结束报文       |          | （前缀）ENT     |
| 放弃报文       | 码组放弃     | （前缀）ENQ     |
|            | 站放弃      | （前缀）ENT     |
| 拆线报文       |          | （前缀）DEL ENT |
| 同步报文       |          | SYN SYN     |

反向监控报文与信息报文传送方向相反，是从站向主站发送的控制序列。主要用于对询问的应答和数据链路的控制。反向监控报文的形式如表 11-3 所示。

表 11-3 反向监控报文

| 名 称  |       | 控 制 报 文             |
|------|-------|---------------------|
| 肯定回答 | 选择报文  | (前缀) ACK 非编号方式应答    |
|      |       | (前缀) DLE0 编号方式应答    |
|      | 信息报文  | (前缀) ACK 非编号方式应答    |
|      |       | (前缀) DLE0 对偶数编号码组应答 |
|      |       | (前缀) DLE1 对奇数编号码组应答 |
| 否定回答 | 选择报文  | (前缀) NAK            |
|      | 信息报文  | (前缀) NAK            |
|      | 探询报文  | (前缀) EOT            |
| 结束请求 | 返回控制态 | (前缀) EOT            |
|      | 返回中性态 | (前缀) EOT            |
| 中断请求 | 码组中断  | (前缀) EOT            |
|      | 站中断   | (前缀) DLE 3/12       |
| 拆线   |       | (前缀) DLE EOT        |

对主站的信息报文和正向监控报文，从站必须应答，有肯定回答（ACK）、否定回答（NAK）等。主站发出结束报文后，若从站同意结束两站之间的通信，回答（前缀）EOT，若从站也有信息需发送给主站，回答（前缀）ENQ。此时主从站的位置相互交换，即原来的从站变成了主站，原来的主站变成了从站。

11.2.3 面向比特型的传输控制规程

在面向比特型的传输控制规程中，ISO 制定的高级数据链路控制规程（HDLC）所覆盖的功能范围最广。下面就以 HDLC 为例，说明面向比特型传输控制规程。

1. HDLC 链路工作方式

因为 HDLC 是为了满足各种应用而设计的，所以考虑了它既能在交换线路上工作，也能在专用线路上工作，既能用于点对点结构，也能用于一点对多点结构；既能采用双向交替方式传输，也能采用双向同时的方式传输。故根据通信双方的链路结构和应答方式，HDLC 为通信操作定义了两种类型，即操作方式和非操作方式。

具体 HDLC 的链路操作有 3 种方式：

- （1）正常响应方式（NRM）：从站只有得到主站探询之后，才能传送有关帧；
- （2）异步响应方式（ARM）：从站不必要得到主站的探询，就可自动地传送有关帧；
- （3）异步平衡方式（ABM）：链路两段都为组合站，任何一站在任意时刻都可发送有关帧命令，无须对方的许可。

HDLC 的非操作方式也有 3 种，说明如下：

- （1）正常断开方式（NDM）：处于这种方式时，从站在逻辑上与数据链路断开，不能发

送和接收信息；

(2) 异步断开方式 (ADM)：处于这种方式时，从站和组合站在逻辑上与数据链路断开，不允许发送信息，但从站或作为命令接收器的组合站具有异步响应方式的机会；

(3) 初始化方式 (IM)：在此方式中，从站和组合站的数据链路控制程序可以分别通过主站或另一个组合站进行初始化或重新生成。

2. HDLC 规程类别

相应于 HDLC 的 3 种操作方式 (NRM、ARM 和 ABM)，有 3 种规程类别，它们是：不平衡操作的正常响应类别 (UNC)、不平衡操作的异步响应类别 (UAC) 和平衡操作的异步响应类别 (BAC)。

HDLC 规程对上述 3 个基本规程定义了其命令/响应子集。另外为了满足特定系统的应用要求，又规定了若干可选择的功能。在通信双方不了解对方使用什么规程类别和哪些可选功能时，可先通过使用 XID 命令 (参见表 11-4) 进行交换协商。

从链路访问规程来看，HDLC 分为链路访问规程 (LAP)、平衡型链路访问规程 (LAPB)、ISDN 的 D 信道链路访问规程 (LAPD)。

3. HDLC 数据链路信道状态

数据链路信道状态分为工作和空闲两种。链路处于工作状态时，主站、从站或组合站发送的是一个帧，或者是连续发送 7 个“1”放弃某帧，或者在帧之间连续发送帧标志序列来填充；链路处于空闲状态时，可在一个站检测出至少 15 个连“1”，表明远端站已停止发送数据。

4. HDLC 帧结构

在 HDLC 中，无论是信息报文还是控制报文都是以帧作为基本单位的，必须符合帧的格式，HDLC 帧的格式如图 11-4 所示。

在 HDLC 帧结构中含有标志字段 F、地址字段 A、控制字段 C、信息字段 I 和帧校验序列字段 FCS 5 个字段。

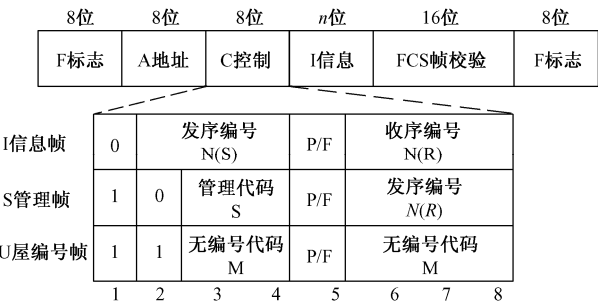


图 11-4 HDLC 帧格式

1) 标志字段 F

标志字段用来表示一帧的开始和结束，用于帧的同步。HDLC 规程指定采用 6 个连“1”的 8 bit 序列 01111110 为标志字段。

在一个帧的中间（F~F）的任何位置不允许出现和标志相同的比特序列，为此，HDLC 规程采用“0”插入技术来实现透明性传输。“0”插入技术可用图 11-5 说明。

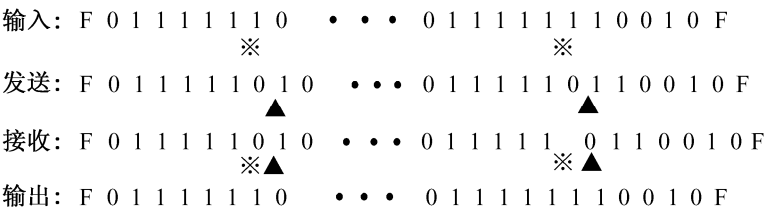


图 11-5 “0”插入与删除技术示意图

当发送站发现有 5 个连“1”的非标志序列，在第 5 个“1”后面自动插入一个“0”。接收端检查到非标志序列的其他比特序列中有 5 个连“1”，自动把后面“0”删去。“0”插入技术的引入，使得 HDLC 帧所传送数据达到透明传送。

2) 地址字段 A

地址字段用来写入发送站和接收站的地址。对于命令帧，该字段为对方的站地址；对于响应帧，该字段指出的是发送响应的数据站地址。

通常地址段有 8 位，其 256 种组合，表示 256 种编址。在 HDLC 规程中，约定全“1”比特为全站地址，用于对全部站点的探询，全“0”比特为无站地址，用于测试数据链路的工作状态。

当站的数目超过 256 个时，可进行地址扩展。扩展方法为 8 位地址的最低位置“0”，表示后面紧跟的 8 位也是地址组成部分，当最低位置“1”，说明后续字节不是扩展地址。使用地址扩展后，单个 8 bit 的地址范围就变成了 128 个。

3) 控制字段 C

控制字段用来表明帧类型、帧编号、命令和控制信息。

C 字段为 8 bit，也允许扩充到 16 bit。

按帧功能可将 C 字段分成信息帧（I 帧）、监控帧（S 帧）和无编号帧（U 帧）3 种，如图 11-6 所示。各帧作用如下。

（1）信息帧 I。控制字段 C 的第一个比特为 0 表示信息帧，用来传输终端用户的信息，并对收到的数据进行确认，另外还有执行探询命令的功能。这种帧格式发送时，含有发送序号 N(S)，指明当前发送帧的编号，并有命令的含义。

帧中含有另一序号为接收序号 N(R)，指明接收站所期待接收的帧号并确认前面各个帧，具有应答意义。例如，N(R)的值设置为 4，收到该字段帧的数据站（主站、次站或组合站）就确认，对方发送的 0、1、2、3 号帧都已被正确接收，而和它通信的那个站正等待 N(S)为 4 的帧。一般在正常操作中，N(S)和 N(R)都为 3 bit，因此表示的最大帧序号为 7。

（2）监控帧 S。控制字段的第 1、2 比特为 10 表示 C 字段此时为监控帧或管理帧。该帧使用时不含信息字段，仅是用于执行数据链路的监控功能，如确认、要求重发和请求暂停等。

根据该帧第 3、4 比特的组合可定义 4 种不同管理功能，它们所表示的意义见表 11-4 所示。监控帧中有接收序号 N(R)，而无发送序号 N(S)。



另外， $N(R)$ 的含义随 S 帧类型的不同而不同，它可以是命令帧，也可以是响应帧。

(3) 无编号帧 U。控制字段的第 1、2 比特为 11 表示 C 字段此时为无编号帧。无编号帧不含有信息字段，也不包含任何用于提供工作方式设置、链路建立、链路拆除等附加的链路控制功能。由于帧中无顺序号，故称为无编号帧。它含有 5 个修饰比特的 M，最多可定义 32 个附加控制功能。常用附加控制功能如表 11-4 所示。

表 11-4 HDLC 命令与应答一览表

| 格 式 | 命 令                | 响 应           | 编 码 |   |      |   |     |   |      |   |
|-----|--------------------|---------------|-----|---|------|---|-----|---|------|---|
|     |                    |               | 1   | 2 | 3    | 4 | 5   | 6 | 7    | 8 |
| I 帧 | I (信息)             | I (信息)        | 0   |   | N(S) | - | P/F |   | N(R) | - |
| S 帧 | RR (接收准备好)         | RR (接收准备好)    | 1   | 0 | 0    | 0 | P/F |   | N(R) | - |
|     | REJ (拒绝)           | REJ (拒绝)      | 1   | 0 | 0    | 1 | P/F |   | N(R) | - |
|     | RNR (接收未准备好)       | RNR (接收未准备好)  | 1   | 0 | 1    | 0 | P/F |   | N(R) | - |
|     | SREJ (选择拒绝)        | SREJ (选择拒绝)   | 1   | 0 | 1    | 1 | P/F |   | N(R) | - |
| U 帧 | SNRM (置正常响应方式)     | RD (请求断开)     | 1   | 1 | 0    | 0 | P/F | 0 | 0    | 1 |
|     | SARM (置异步响应方式)     |               | 1   | 1 | 1    | 1 | P/F | 0 | 0    | 0 |
|     | SABM (置异步平衡方式)     |               | 1   | 1 | 1    | 1 | P/F | 1 | 0    | 0 |
|     | SNRME (置扩充的正常响应方式) |               | 1   | 1 | 1    | 1 | P/F | 0 | 1    | 1 |
|     | SARME (置扩充的异步响应方式) |               | 1   | 1 | 1    | 1 | P/F | 0 | 1    | 0 |
|     | SABME (置扩充的异步平衡方式) |               | 1   | 1 | 1    | 1 | P/F | 1 | 1    | 0 |
|     | SIM (置初始化方式)       | RIM (请求初始化方式) | 1   | 1 | 1    | 0 | P/F | 0 | 0    | 0 |
|     | DISC (断开)          | DM (断开方式)     | 1   | 1 | 0    | 0 | P/F | 0 | 1    | 0 |
|     | UP (无编号探测)         | UI (无编号信息)    | 1   | 1 | 0    | 0 | P/F | 1 | 0    | 0 |
|     | UI (无编号信息)         |               | 1   | 1 | 0    | 0 | P/F | 0 | 0    | 0 |
|     | XID (交换标志)         |               | 1   | 1 | 1    | 1 | P/F | 1 | 0    | 1 |
|     | REST (复位)          |               | 1   | 1 | 1    | 1 | P/F | 0 | 0    | 1 |
|     | TEST (测试)          | TEST (测试)     | 1   | 1 | 0    | 0 | P/F | 1 | 1    | 1 |
|     |                    | UA (无编号确认)    | 1   | 1 | 0    | 0 | P/F | 1 | 1    | 0 |
|     |                    | FRMR (帧拒绝)    | 1   | 0 | P/F  | 0 | 0   | 1 |      |   |

(4) 探测/终止 (P/F) 位。在以上 3 种帧中，均有 1 bit 的 P/F 位，意为命令的探测/或响应的终止，只有置“1”时才有用。在信息（命令和响应）帧中， $N(S)$ 、 $N(R)$ 和 P/F 的功能是相互独立的；在监控（命令和响应）帧中  $N(R)$ 和 P/F 的功能是相互独立的。

P/F 被主站和从站用来实现如下 4 种功能：

- ① P/F 位在被主站使用时称为 P 位，在被从站使用时称为 F 位。在一条链路上，任何时刻都只能有一个 P 位可以存在（等待 F 位响应）。一般 P 位置“1”可以用做检验点，用来探测从站的状态。检验点在各种形式的自动化中很为重要，它是机器用来清除二义性，清理以前堆积的事务的一种技术。
- ② 主站用 P 位请求从站发一个状态响应；P 位还可以表示探测，用于授权或启动从站传输信息。

③ 从站对主站 P 位的响应, 根据主站命令的不同, 可为信息帧、监控帧或无编号帧, 且带 F 位表示从站的响应。另外, 在正常响应方式时, F 位还可表示从站发送数据的末位, F 置“1”, 预告传输即将结束。

④ P/F 位可以以几种方式来使用和解释:

- 在 NRM 方式中, 从站只有在收到一个 P 位为“1”的帧后才可以发信。主站要求从站发送信息帧 (I) 时, 可以发一个 P 位为“1”的帧, 或者发某些监控命令帧 (S), 如 RR、REJ 或 SREJ 等, 并使 P=1。
- 在 ARM 方式和 ABM 方式中, 不收到 P 位为“1”的命令也可以发送信息帧。这时 P 位为“1”用来请求尽快发一个 F 位为“1”的响应。
- 在 ARM 方式和 ABM 方式中, 接到 P 位为“1”的命令后发送一个 F 位为“1”的“准备接收 (RR)”帧, (这实际是非抢占优先权问题)。在双向同时 (全双工) 发送时, 当从站正在发信时收到一个 P 位为“1”的命令, 那么在它后面最早的那个响应中设置 F 位为“1”; 发送 F 位为“1”的帧并不要求从站停止发信。在发了 F 位为“1”的帧后面还可以继续发信。可见在 ARM 方式和 ABM 方式中, 从站并不把 F 位解释成发信的结尾, 它只是解释成对前面那个帧的响应。

#### 4) 信息字段 I

信息字段用于传送用户数据。信息字段的长度是可变的, 理论上不限长度。实际应用中, 其长度由收发站的缓冲器的大小和线路的差错情况决定, 通常不大于 256 字节。

#### 5) 帧校验序列字段 FCS

帧校验字段用于对帧进行循环冗余校验, 校验的范围包括除标志字段之外的所有字段 (A、C、I 字段) 的内容, 但为了进行透明传输而插入的“0”不在校验范围内。该字段一般为 16 bit, CRC 码校验时的生成多项式为  $g(x)=x^{16}+x^{12}+x^5+1$ 。为了满足更高的要求, 也可采用 32 bit 的 CRC 码校验。

### 5. 异常状态的报告和恢复

链路异常是指传输出现差错、数据站故障或误操作等情况。

链路异常时应执行恢复规程, 使链路恢复正常, 常用的恢复措施有以下几项:

(1) 忙。接收站由于内部缓冲器限制等原因暂时不能接收信息时, 则称其处在“忙”状态, 这时可通过发送 RNR 帧通知对方。“忙”状态的消除可以通过发送 RR, REJ, SREJ, UA 以及置工作方式命令等来进行, 并可接收 I 帧。

(2) N(S) 差错。当收到 I 帧无校验 FCS 差错, 但该帧的 N(S) 不等于预期所要接收的序号时, 则表明 N(S) 顺序出现了差错情况, 处于异常链路工作状态。根据差错出现的不同原因, 响应采取 P/F 检验恢复, REJ 恢复, SREJ 恢复和超时恢复等措施。

(3) FCS 校验差错。当接收器发现某帧有校验差错, 将其丢弃。

(4) 命令/响应被拒绝。当收到的帧无差错, 但其控制字符包含了未定义或不能实现的命令或响应以及帧格式无效和信息字段长度超过容许的最大长度时, 链路处于命令/响应拒绝状态。通常采用置工作方式命令去恢复。

(5) 竞争状态。当同时有多个站去占用链路时, 链路处于竞争状态。常采用系统规定的

超时条件去恢复。

6. HDLC 传输过程举例

例 11-1 从站发送信息的正常响应方式 (NRM) 的双向交替传输数据链路的建立和数据传输。该过程如图 11-6 所示。

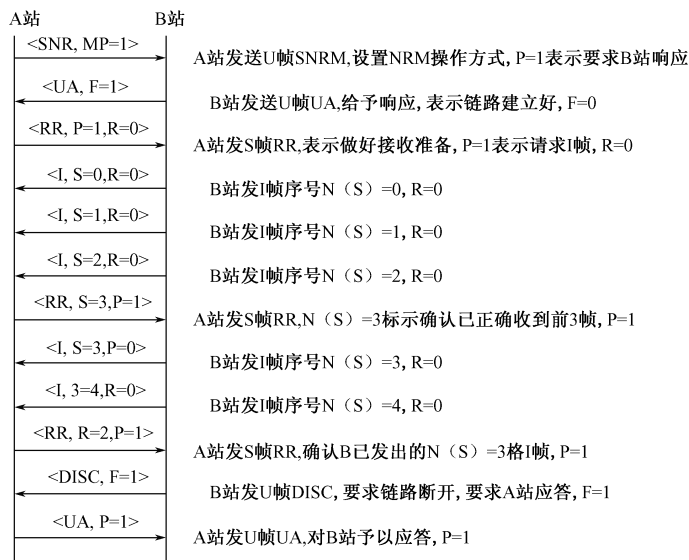


图 11-6 从站发送信息的 NRM 方式下数据传输过程

在初始状态,主站 A 向从站 B 发送 SNRM 命令,以设置 NRM 工作方式, P=1 表示主站发送的是命令,要求从站响应。

从站收到命令后发送 UA 帧以示确认,这时数据链路已建立完毕, F=1 表示是从站送出的命令,要求主站响应。

主站为了要求从站发送数据信息,发送 RR 命令,表示主站已准备好接收数据信息, P=1 以示允许从站发送数据。所以从站接收到 RR 命令后开始发送 I 帧,帧序号用 N(S)表示。

当主站连续收到从站发送 I 帧的第 3 帧(N(S)=2),送出 RR 命令(P=1)以表示从站可以继续发送 I 帧,同时对前面收到的 3 帧进行确认(N(R)=3)。

从站在数据发送完毕时,在最后一个 I 帧置 F=1。

如果上面的数据链路中主站要向从站发送数据信息,则主站在收到从站送来的 UA 响应帧后,就可以马上发送 I 帧了。

例 11-2 NRM 方式双向同时传输数据链路的数据传输。

传输过程如图 11-7 所示。它适合于全双工通信。

数据链路的建立过程与例 11-1 相同,但主站在送出第 1 个 I 帧时, P=1 表示同时允许从站发送 I 帧,双向同时传输数据信息。但传输 I 帧的同时,对收到的 I 帧进行确认。

例 11-3 退回 N 步的差错恢复过程。

图 11-8 给出了退回 N 步的差错恢复过程。当 A 站发送编号 7 的 I 帧输出错,由接收站 B 站检测出出错情况, B 站等待 A 站的询问,当收到 P=1 的询问帧时, B 站用 RR 帧响应,并

用  $N(R)=7$  要求 A 站重发编号为 7 以后的 I 帧，从而达到差错恢复。这种方式适用于双向交替方式工作的链路，对于双向同时方式工作的链路是不适用的。

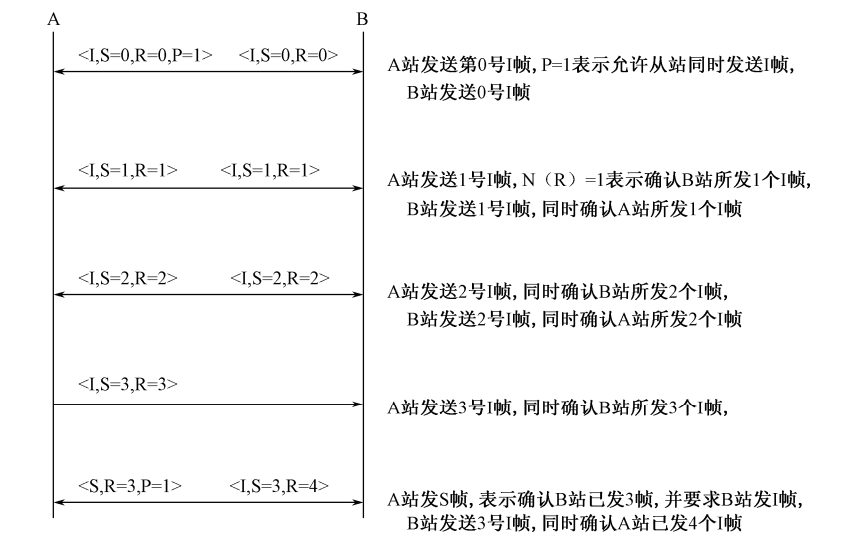


图 11-7 NRM 方式双向同时数据传输传输过程

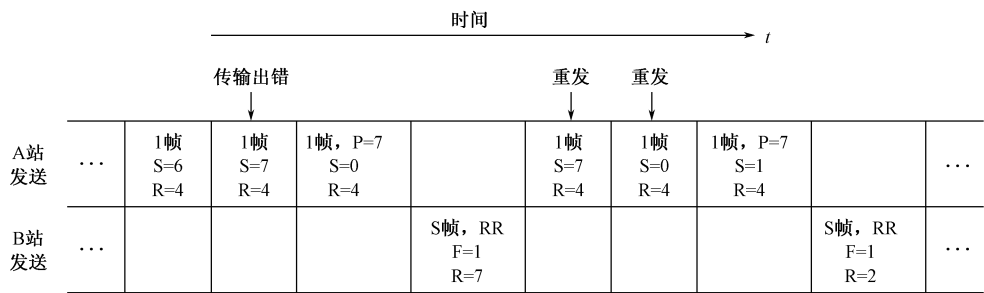


图 11-8 退回 N 步的差错恢复过程

例 11-4 用 SREJ 帧进行差错恢复过程。

图 11-9 给出了差错恢复过程。用 SREJ 帧来指出出错的帧号，发站 A 重发出错的帧，而对于已发送的未出错的帧不予重发，从而提高传输效率。

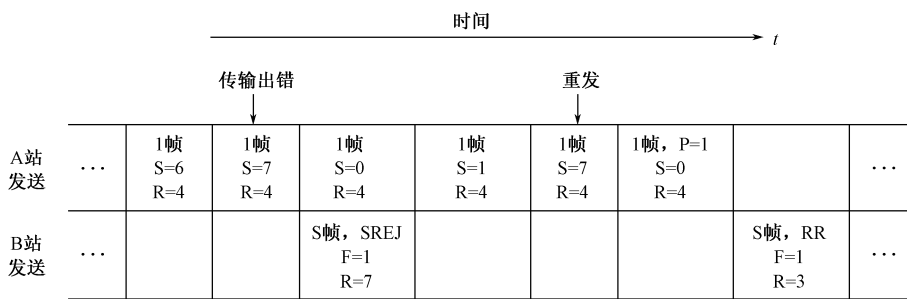


图 11-9 SREJ 帧进行差错恢复过程

## 7. 数据链路传输控制规程比较

与字符型控制规程相比较，HDLC 规程有以下特点：

(1) 透明传输。HDLC 规程对所传输的数据信息的比特序列的结构没有限制，是“0”插入技术来实现。面向字符型的传输控制规程靠转义字符 DLE，它降低了传输效率，增加了处理的复杂性。

(2) 可靠性高。在 HDLC 规程中，差错校验的范围为整个帧（除了 F 标志之外）。

(3) 传输效率高。因为 HDLC 的 I 帧的编号传输，在通信中可以不必等待对方响应进行确认后才传输新的数据信息，而且也不必对每个 I 帧都予以确认；在差错时，可以采用选择重发的方法。这些都大大提高了工作效率。另外，从帧结构的角度来看，帧内用于寻址和控制所占的比例也大大小于字符型传输规程，这从另一个方面说明 HDLC 的传输效率较高。

(4) 应用广泛，适应力强。从工作方式上看，HDLC 能选择采用双向交替和双向同时工作两种工作方式；从它所适用的数据电路来看，它既能用于二线电路，也能用于四线电路；既能用于交换线路，又能用于专用线路；它既适应点对点方式工作，也能适应一点对多点方式工作；它既能用于电路交换的数据通信网，也能用于分组交换的数据通信网，它还能用于局域网和卫星通信的链路控制。

(5) 结构灵活。在 HDLC 中，传输控制功能和处理功能分离，层次清楚，应用非常灵活。尽管 HDLC 规程要比面向字符型的传输控制规程要复杂些，但由于计算机、单片机、微处理器、VLSI 的发展，出现了许多专用或多用的面向比特的传输控制集成电路芯片，如 Z8-SIO、Intel 8274 和 MC68000 等，所以实现该规程极为方便。

## 11.3 无线数据通信网中的信道接入协议

信道接入协议，即媒体访问控制（MAC），是无线数据通信协议的重要组成部分。目前在无线数据网中常用的信道接入协议可分为：单信道接入协议与多信道接入协议。

### 11.3.1 单信道无线接入协议

单信道接入是无线数据通信网中常用的接入形式。在单信道条件下，节点的发送和接收只占用一个信道，工作于半双工方式。常用的单信道基本接入协议主要有 ALOHA、CSMA/CA、MACA 等。

#### 1. ALOHA 协议

ALOHA 协议是适用于在单信道条件下使用的一种随机接入协议。ALOHA 协议描述为：网络中的所有节点共享信道，任何节点需要发送数据时，不需要获取信道工作状态，随时发送，检测到碰撞，给每个用户分配不同的重发间隔时间，重新发送。ALOHA 协议简单，吞吐率较低，最大值只能达到 18.4 %。因此它只适用于节点数较少、业务量较低的网络中使用。

时隙 ALOHA 协议（S-ALOHA）是对传统的 ALOHA 协议的改进，它将整个信道的占用时间分成长度为 T 的时隙，并规定节点发送数据的起始时刻必须与划分的时隙起点对齐，即只能在每个时隙开始时刻发送一个数据帧。时隙 ALOHA 与纯 ALOHA 相比，吞吐率提高了

1 倍，达到了 36.8 %，但付出的代价是全网需要严格的时间同步。

在时隙 ALOHA 的基础上，如果将时隙分给固定的节点使用，则构成了 TDMA 协议。TDMA 协议有效避免了信道数据碰撞，但 TDMA 需要精确的时间同步，且信道利用率不高。

码分多址（CDMA）方式能够支持多个用户同时共享一个时间资源和频带资源，结合了时隙 ALOHA 和 CDMA 两种技术的 MAC 方案是一另种改进的时隙 ALOHA 方案，由于 ALOHA-CDMA 方案，采用了 CDMA 技术，利用了时隙时间维度、码维度的二维信息，它可在一个时隙中支持更多个网络节点同时接入。其接入效率受到码长、时隙数目、节点数目三者的联合制约，如果能估计节点数目，就可以指导码长、时隙数目的合理安排，以便达到最佳接入效率。

## 2. CSMA/CA 协议

CSMA/CA(载波监听/碰撞避免协议) 是指节点在发送数据时，先监听信道的工作状态，如果信道已经被其他节点占用，则延迟发送以避免数据碰撞。

由于 CSMA/CA 中在发送数据前先监听了信道工作状态，因此可以降低信道碰撞概率，但同时也增大了发送延时。

## 3. MACA 协议

MACA 协议（多址接入碰撞避免协议）包括 RTS、CTS、DATA 和 ACK 四个过程。但它充分利用了 RTS 与 CTS 帧，避免了隐终端和暴露终端的问题。

MACA 协议规定，当两个节点 A、B 进行 RTS-CTS 握手时，将周围的其他节点分成 3 类：

第 1 类是既能监听到 RTS 帧，也能监听到 CTS 帧的节点。显然这类节点处于发送节点和接收节点的通信覆盖范围之内，可以监听到整个数据通信过程，因此在整个通信过程中，应该保持静默；

第 2 类节点是只能监听到 RTS 帧的节点，这类节点只处于发送节点的通信覆盖范围之内。在监听到 RTS 帧后的一段时间内保持静默，以确保发送节点能正确地收到 CTS 帧。当发送节点开始发送数据帧后，这类节点就不必保持静默，因为它处于接收节点通信范围之外，不会对接收节点产生干扰；

第 3 类节点是只能监听到 CTS 帧的节点，这类节点在监听到 CTS 帧后，在一段足够长的时间内保持静默，以避免干扰接收节点。

# 11.3.2 多信道接入协议

多信道接入协议是指利用 2 个以上的信道进行数据传输。一般来说，多信道协议将可用信道划分为 2 类：一类是控制信道，用于传帧长较短的控制帧；另一类是数据信道，用来传输帧长较长的数据帧。常用的多信道传输协议主要有：BTMA、DBTAM、扩频码的信道接入协议等。

## 1. BTMA 接入协议

BTMA（基于忙音的多信道接入协议）将信道分为控制信道和数据信道。控制信道用于传送 RTS 帧、CTS 帧和忙音。协议工作过程如图 11-10 所示。

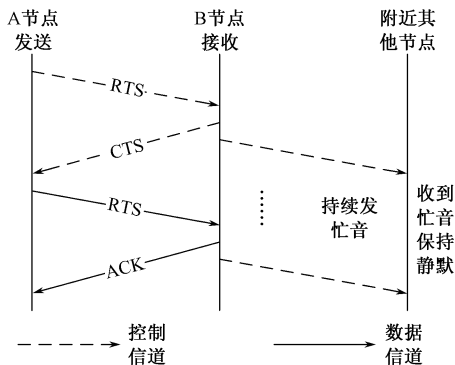


图 11-10 BTMA 协议示意图

当 A 节点有数据向 B 节点发送时，A 节点先在控制信道上发送 RTS 帧，B 节点在收到 RTS 帧后，先在控制信道上响应 CTS 帧，然后在控制信道上发送一个忙音，表示自己正处于接收数据状态，不希望其他节点在数据信道上干扰。

接下来 A 节点开始在数据信道上发送 DATA 帧。处于 B 节点通信覆盖范围内的其他节点收到忙音后，立即在数据信道上保持静默，以免干扰 B 节点，直至 B 节点完成数据接收为止。

从图 11-10 可以看出，基于 BTMA 协议较为复杂，它可以有效地避免在接收数据时发生隐终端问题，但要实现 BTMA 协议，每个节点需要 2 个信道，且 2 个信道要能同时工作，这对接入设备要求较高。

## 2. 基于扩频码的信道接入协议

基于扩频码的信道接入协议是利用扩频码的正交特性，使得网络中各节点可以同时占用信道而互不干扰。

### 1) 基于接收端的扩频码接入协议

网络中的每个节点都具有一个唯一的标识 (ID)，并且被分配一个唯一的、正交的扩频码。当一个节点需要向另一个节点发送数据时，它先在信息数据中添加上接入节点的 ID 值，然后使用接收节点所对应的扩频码对发送的信息数据进行调制。在接收端，节点使用自身的扩频码对接收到的信号解调，并根据其中的 ID 值决定是否接收该信息。使用该协议时，节点的接收设备较为简单，只要在系统干扰容限之内，系统可以允许不同的节点可以同时占用信道。当 2 个不同的节点同时向另一个节点发送信息时，在接收端可能产生干扰。

### 2) 基于发送端的扩频码接入协议

网络中的每个节点都被具有一个唯一的标识 (ID)，并且分配了一个唯一的、正交的扩频码。当某一节点向另一个节点发送数据时，它先在信息数据中添加上接入节点的 ID，然后本端的扩频码对发送的信息数据进行调制。在接收端，接收节点使用系统中所有可能的扩频码对接收信号进行解调，然后根据其中的 ID 值决定是否接入信息。在该协议中，当 2 个节点同时向另一个节点发送信息时，接入节点仍然能够将两个信息区分开来。但要实现这种协议，显然要求节点具有多个扩频信道同时解调的能力，接入设备的复杂程度较高；

3) 基于公共扩频码的接入协议

基于公共扩频码的接入协议是指系统使用若干个公用的扩频码，这些公用的扩频码经过优选，相互之间有十分优异的正交性。当网络中的某个节点需要发送数据，先通过监听，在所有公共扩频码中寻找一个尚未被使用的扩频码，然后再用该扩频对信息数据进行调制。接收端在使用所有可能的公共扩频码对接收到的信号解扩，根据信息中包括的目 ID 决定是否接收该信息。基于公共扩频码的接入协议要求接收机具有多个扩频信道接收能力，但信道数量比基于发送端的扩频码接入协议少。

从上述分析可知，由于扩频具有良好的正交性，可以容忍网络中的一些数据碰撞，因此隐终端问题和暴露终端问题可以得到较好地克服。与单信道接入协议和多信道接入协议相比，基于扩频码的信道接入协议可较好地解决不同节点之间争用信道的问题，提高系统的效率，但是它占用的频谱资源也较多，接入设备复杂性较高。

11.3.3 无线信道的隐终端和暴露终端问题

从网络通信的角度来看，无线信道具有开放性、共享性的特点。开放性意味着在通信距离范围内的所有节点都具有获取信息资源的能力，而共享性意味着网络中的所有节点都具有占用信道带宽、发送信息的权限。无线信道的开放性和共享性会导致隐终端和暴露终端问题。

1. 隐终端

隐终端问题如图 11-11 (a) 所示，A、B、C、D 共 4 个点分布位置如下，B 节点在 A 节点的通信覆盖范围之内，C 节点在 A 的覆盖范围之外。当 A 节点有数据需要发送给 B 节点，A 先向 B 发送 RTS 帧。

由于 C 节点不在 A 节点的通信覆盖范围之内，不可能监听到 A 节点发出的 RTS 帧。所以，此时如果 C 节点向 D 节点发送数据，就会干扰 B 节点对 RTS 帧的接收。这种情况下，称 C 节点为隐终端。

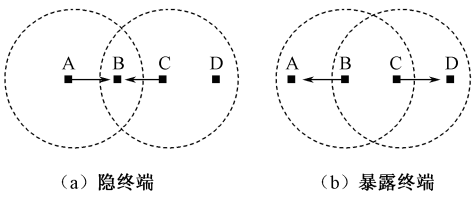


图 11-11 隐终端和暴露终端问题

2. 暴露终端

暴露终端问题如图 11-11 (b) 所示，A 节点处于 B 节点通信覆盖范围之内，处于 C 节点通信覆盖范围之外，D 节点处于 C 节点通信覆盖之内，但处于 B 节点通信覆盖范围之外。在理想情况下，如果 B 节点向 A 节点发送数据时，C 节点也可以向 D 节点发送数据，此时 A 节点应该能正确收到 B 节点发来的数据，因为 A 节点远离 C 节点，它不会受到 C 节点的干扰；同理，D 节点也应该能收到 C 节点发来的数据，因为 D 节点远离 B 节点，它不会受到 B



节点的干扰。但是在某些协议中，B、C 节点却不能同时发送数据。例如在 CSMA/CA 协议中，当 C 节点听到了 B 节点正在发送数据，就会启动避让规则，暂缓向 D 节点发送数据，这就增加了 C 节点到 D 节点的数据传输时延。在这种情况下，称 C 节点为暴露终端。

从以上分析可知，由于无线信道的特点和信道接入协议在设计上存在的缺陷，导致了隐终端和暴露终端问题，降低了网络的效率。

## 11.4 无线局域网协议

### 11.4.1 无线局域网解决方案与运作模式

无线局域网解决方案主要有：

(1) 无线个人网 (WPAN)：主要用于个人用户工作空间，典型距离复盖几米，可以与计算机同步传输文件，访问本地外围设备（如打印机等）。目前主要技术包括“蓝牙”和红外 (IrDA)。

(2) 无线局域网 (WLAN)：主要用于宽带家庭、大楼内部以及园区内部，典型距离复盖几十米至上百米。

(3) 无线 LAN-to-LAN 网桥：主要用于大楼之间的联网通信，典型距离为几公里。

(4) 无线城域网 (WMAN) 和无线广域网 (WWAN)：覆盖城域和广域环境，主要用于 Internet/email 访问，但提供的带宽比无线局域网要低很多。

一般架设无线网络的基本配备就是无线网卡及一台 AP (Access Point, 无线接入点或桥接器)，如此便能以无线的模式，配合既有的有线架构来分享网络资源，如果只是几台电脑的对等网，也可不要 AP，只需要每台电脑配备无线网卡。使用 AP 在空间中制造一个无线网络区域，网络内的无线网卡搜索到无线网络信号，进行参数修正以后就可以连接到无线网络中。

无线局域网标准目前主要技术为 IEEE 802.11 系列。IEEE 802.11 定义了两种运作模式：特殊 (Ad Hoc) 模式和基础 (Infrastructure) 模式。

在 Ad Hoc 模式下，Ad Hoc 模式用于不提供无线 AP，无线客户端直接相互通信。使用 Ad Hoc 模式通信的两个或多个无线客户端就形成了一个独立基础服务集。

在 Infrastructure 模式下，至少存在一个无线 AP 和一个无线客户端。无线客户端使用无线 AP 访问有线网络的资源。支持一个或多个无线客户端的单个无线 AP 称为一个基础服务集 (Basic Service Set, BSS)，BSS 是无线局域网的基本单元，即前面说的小区，在一个基本服务集 BSS 内，所有的站均运行同样的 MAC 协议并以争用方式共享同样的媒体。

一个 BSS 可以通过 AP 连接到主干分配系统 DS (Distribution System)。DS 相当于一个有线主干局域网。两个或更多的基本服务集组成扩展服务集 (Extended Service Set, ESS)，而这些基本服务集通过分配系统连接在一起，如图 11-12 所示。一个 ESS 是单个逻辑网段（也称为一个子网），并通过它的服务集标识符 (Service Set Identifier, SSID) 来识别。如果某个 ESS 中的无线 AP 的可用物理区域相互重叠，那么无线客户端就可以漫游，或从一个位置（一个无线 AP）移动到另一个位置（另一个不同的 AP），同时保持网络层的连接。

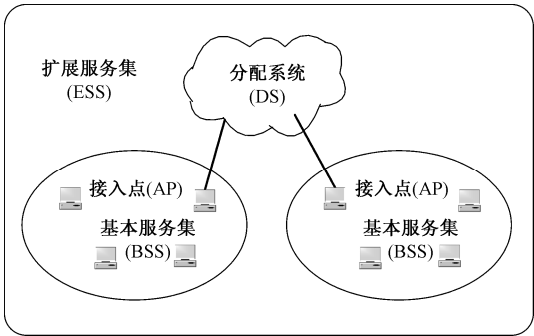


图 11-12 IEEE 802.11 的基本服务集和扩展服务集

11.4.2 IEEE 802.11 标准中的物理层

目前 802.11 中制订了多种介质的实体，对未来技术的扩充性，也都提供了多重速率的功能。下面介绍几种常用的实体。

- (1) IEEE 802.11a 工作在 5 GHz 频段、最大速率可达 54 Mb/s、采用 OFDM 调制技术，采用 802.11a 标准的 WLAN 可以同时支持多个相互不干扰的高速 WLAN 用户。
- (2) IEEE 802.11b 使用直接序列扩频 DSSS 调制技术，在 2.4 GHz 频带实现 11 Mb/s 速率的无线传输。由于 DSSS 技术的实现比正交频分复用 OFDM 容易，802.11b 已成为当今 WLAN 的主流标准。其最大特点是可以根据无线信道状况的变化，在 1 Mb/s、5.5 Mb/s、11 Mb/s 之间进行速率的动态调整。
- (3) IEEE 802.11g 在 2.4 GHz 频段使用 OFDM 调制技术，使数据传输速率提高到 20 Mb/s 以上；802.11g 标准既能提供与 802.11a 相同的传输速率，又能与已有的 802.11b 设备后向兼容。在速率不大于 11 Mb/s 时，仍采用 DSSS 调制技术；当传输速率高于 11 Mb/s 时，则采用调制效率更高的 OFDM 调制技术。

11.4.3 IEEE 802.11 标准中的 MAC 子层

在网络标准内，各种传输介质的物理层对应到相对的 MAC 层，例如，以太网 MAC 层定义为 802.3，无线局域网 MAC 定义为 802.11，各种不同传输介质的物理层对应的不同的 MAC 层，如图 11-13 所示。



图 11-13 不同传输介质的物理层对应的 MAC 层

由于以太网 (Ethernet) 成为现存局域网络结构的绝大多形式，CSMA/CD (Carrier Sense Multiple Access/Collision Detection) 载波监测多址接入协议也成为局域网采用最多的 MAC 协

议。CSMA/CD 适宜于总线型局域网拓扑结构的随机竞争型媒体访问控制。总线型网络允许同一时刻只有一个节点（Node）发送数据，一旦两个或以上节点同时发送数据，则会发生数据碰撞，数据不能正常发送和接收。CSMA/CD 协议就是尽可能保证网络上同时只有一个节点发送数据，减小数据“碰撞”概率。

为了尽量减少碰撞，IEEE 802.11 标准设计了独特的 MAC 子层，包括分布式协调控制功能（Distributed Coordination Function, DCF）和点协调控制功能（Point Coordination Function, PCF），PCF 也叫中心控制、集中控制。协调控制功能是决定在 BSS 内工作的一个站，通过无线媒体何时允许发送和可能接收协议单元的逻辑功能。

DCF 是 IEEE 802.11 MAC 协议最基本的媒体接入方法，作用于基本服务群和基本网络结构中，可在所有站实现，主要采用 CSMA/CA 的分布式媒体接入方法，让各个站通过争用信道来获取发送权。DCF 向上提供争用服务。

PCF 使用集中控制的接入算法（一般在接入点实现集中控制），用类似于轮询的方法将发送数据权轮流交给各个站，从而避免了碰撞的产生。对于时间敏感的业务，如分组语音，应当使用点协调功能 PCF，可支持无争用型业务。其基本结构如图 11-14 所示。

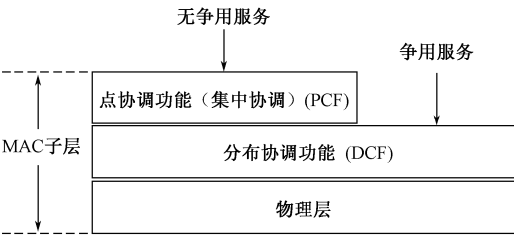


图 11-14 IEEE 802.11 标准中的 MAC 子层基本结构

11.4.4 载波侦听多址访问/碰撞回避机制

载波侦听多址访问/碰撞回避机制，即 CSMA/CA（Carrier Sense Multiple Access/Collision Avoidance）机制。

IEEE 802.3 总线型局域网（有线以太网）在 MAC 层的标准协议是 CSMA/CD（Carrier Sense Multiple Access/Collision Detection）。所有的以太网，不论其速度或帧类型是什么，都使用 CSMA/CD。图 11-15 描绘了 CSMA/CD 的工作过程。

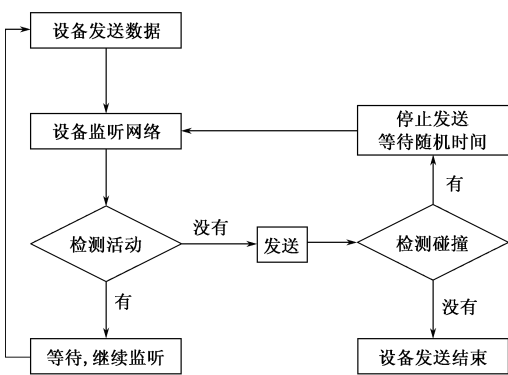


图 11-15 CSMA/CD 的工作过程

因为无线设备不能同时发射和接收，所以 IEEE 802.11 使用碰撞回避策略，而不是如 IEEE 802.3 使用碰撞检测（Collision Detection，CD）。在一个 WLAN 中，不是所有的无线设备都能够直接通信。因此，IEEE 802.11 采用网络分配矢量（NAV）。NAV 是表示媒介空闲剩余时间的值。每个站点的 NAV 都是从媒介传输的帧里取出时间长度值来保持最新值。站点则通过检查 NAV 决定是否发送。有可能 NAV 表示媒介忙，而物理载波检测却显示媒介空闲，这时站点不能发送。因此，NAV 也被称为虚拟载波检测。通过物理载波检测和虚拟载波检测策略的结合，MAC 得以实现 CSMA/CA 的碰撞避免机制。图 11-16 描绘了 CSMA/CA 的工作过程。

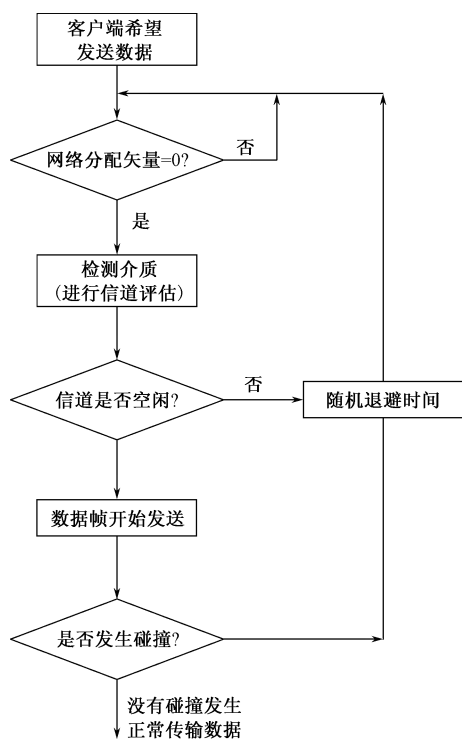


图 11-16 CSMA/CA 的工作过程

从 CSMA/CA 的工作过程知道，查看介质是否空闲通过载波侦听，如果源需要发送数据时，它必须检查传输介质是否正在被别人使用，如果介质未被别人使用，就发送数据包；反之，源主机就必须等待一段时间，等到介质空闲后才能够发送。源通过不断发出 RTS（Request To Send，请求发送）包来检查介质的占用情况，在目的主机返回 CTS（Clear To Send，允许发送）包后，源主机就可以发送了。即请求发送/允许发送协议（RTS/CTS）。

11.4.5 RTS/CTS 协议

RTS/CTS 协议相当于一种握手协议，主要用来解决“站点隐蔽”问题。“站点隐蔽”是指用户站 A 和用户站 C 都未侦听，同时将信号发送至 B 引起信号碰撞，导致发送至 B 的信号都丢失了。RTS/CTS 传输原理如图 11-17 所示。

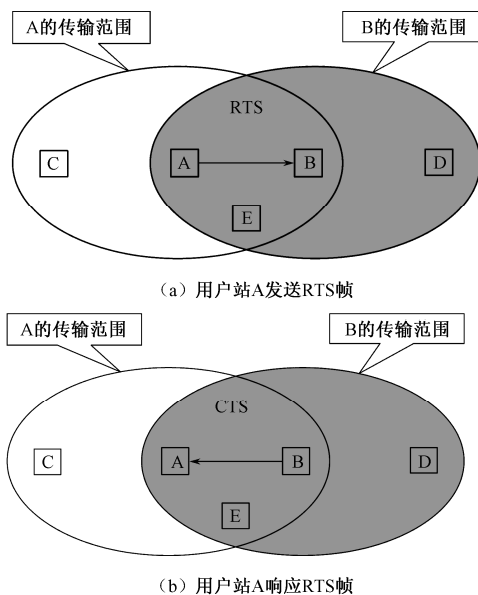


图 11-17 CSMA/CA 协议中的 RTS 和 CTS

图 11-17 (a) 中用户站 A 在向站 B 发送数据帧之前, 先向 B 发送 RTS 帧, 表明 A 要向 B 发送若干数据, 在 RTS 帧中说明将要发送的数据帧的长度。RTS 帧格式如图 11-18 所示。

|     |      |        |        |        |
|-----|------|--------|--------|--------|
| 帧控制 | 持续时间 | 接收地址RA | 发送地址TA | 帧校验FCS |
|-----|------|--------|--------|--------|

图 11-18 RTS 帧格式

图 11-17 (b) 中用户站 B 收到 RTS 帧后, 就向站 A 响应一个允许发送帧 CTS。CTS 帧格式如图 11-19 所示。

|     |      |        |        |
|-----|------|--------|--------|
| 帧控制 | 持续时间 | 接收地址RA | 帧校验FCS |
|-----|------|--------|--------|

图 11-19 CTS 帧格式

在 CTS 帧中也附上 A 欲发送的数据帧的长度 (从 RTS 帧中复制)。A 收到 CTS 帧后就发送其数据帧了。B 接收完数据后, 即向所有用户站广播 ACK 确认帧, 所有用户站又重新可以平等侦听、竞争信道了。

在 A 和 B 两个站附近的一些站 (站 C、D、E) 的反应:

C 在 A 的范围内能收到 RTS, 但收不到 CTS, 因为不在 B 的范围内, D 与 B 相反, 而 E 即能收到 RTS 也能收到 CTS, 因此这 3 个站在整个过程都不影响 A 向 B 发送数据。

发送该帧的目的是将持续时间信息告知邻近站点, 收到 RTS 的站点就用收到的信息更新其 NAV, 从而防止这些站点在被告知时间段内发送信息, 以避免碰撞发生。RTS 帧包含帧控制、持续时间、两个地址和帧校验, 在 RTS 帧中, 持续时间信息是完成一次帧交换过程所需时间。RA 表示无线媒体上的一个站点的 IEEE MAC 地址, RA 标识的站点则发送 RTS 的应答帧 CTS。TA 表示发送该 RTS 帧的站点的地址。FCS 包括一个 IEEE 32 位循环冗余校验码 CRC。

CTS 包含帧控制、持续时间、一个地址和帧校验，该帧用于将持续时间信息告知邻近站点。收到 CTS 的站点就用收到的信息更新其 NAV，从而防止这些站点在被告知的持续时间内发送信息，避免碰撞发生。

CTS 帧是刚刚接收到的 RTS 的响应帧，RA 就是该 RTS 帧中的发送方地址 TA 的拷贝。CTS 帧中的 RA 必须是某个站点的 IEEE MAC 地址。CTS 帧中的持续时间数值也是从刚刚接收到的那个 RTS 中的持续时间域中获得。持续时间以  $\mu s$  为单位。FCS 同 RTS 帧中的 FCS。

使用 RTS 和 CTS 会使整个网络效率下降，但他们相比于数据帧可以忽略，不使用这种控制帧，一旦发生碰撞导致数据帧重发，时间的浪费就更大。

因此，协议 802.11 协议设有 3 种情况供用户选择：

- (1) 当需要传送大容量文件时，使用 RTS 和 CTS 帧；
- (2) 只有当数据帧的长度超过某一数值时，使用 RTS 和 CTS 帧；
- (3) 不使用 RTS 和 CTS 帧。

### 11.4.6 随机的时间等待 IFS

随机的时间等待 IFS (Inter Frame Space) 即帧间间隔。IEEE 802.11 查看介质是否空闲通过载波侦听，而碰撞回避通过随机的时间等待 IFS，碰撞回避使信号发生碰撞的概率减到最小，当介质被侦听到空闲时，优先发送。

先给出只使用 IFS 时的 CSMA 接入算法原理：

(1) 欲发送帧的站先监听信道。若发现信道空闲，则继续监听一段时间 IFS，看信道是否仍为空闲。是，则立即发送数据。信道已经空闲了还继续监听一段时间，是因为 3 种不同数值的 IFS 可将数据划分为不同的优先级，IFS 值小的优先级高。这样能够减少碰撞的机会。

(2) 若发现信道忙，则继续监听信道，直到信道变为空闲。

(3) 一旦信道变为空闲，此站迟延另一个时间 IFS。若信道在时间 IFS 内仍为空闲，则按指数退避算法迟延一段时间。只有当信道一直保持空闲时，该站才能发送数据。这样做可使在网络负荷很重的情况下，发生碰撞的机会大为减小。

常用的几种不同帧间间隔 IFS 如下：

(1) SIFS (short IFS)：SIFS 是最短的 IFS，典型的数值只有  $10 \mu s$ 。当节点获得信道的控制权，为了帧交换序列继续保持信道控制，这时就使用 SIFS，提供了最高优先级。

(2) PIFS (PCF IFS)：即点协调功能 IFS，比 SIFS 长，在 PCF 方式中轮询时使用。仅仅当节点在 PCF 模式下，为了在非竞争周期开始时获得信道的访问控制优先权而使用的。一旦在这个时间内侦听到信道空闲，就可以进行中心控制方式的无竞争的通信。

(3) DIFS (DCF IFS)：即分布协调功能 IFS，节点在 DCF 方式下传输数据帧和管理帧所使用的时间间隔，是最长的 IFS，典型数值为  $50 \mu s$ 。如果载波侦听机制确定在正确接收到帧之后的 DIFS 时间间隔中，信道是空闲的，而且退避时间已经过期，节点将进行发送。

(4) EIFS (Extended IFS)：DCF 方式下用于接收数据错误的情况下的等待时间，为接收节点发送确认帧 (ACK) 提供足够的时间。

以上这些帧间间隔的长度实际上就决定了它们的优先级，即 EIFS 的优先级 < DIFS 的优先级 < PIFS 的优先级 < SIFS 的优先级。当很多节点都在侦听信道时，使用 SIFS 可具有最高的优先级，因为它的时间间隔最短。

图 11-20、图 11-21 说明了帧间间隔的作用。从图 11-20 基本接入方法可看出，当很多站都在监听信道时，使用 SIFS 可具有最高的优先级，因为它的时间间隔最短。

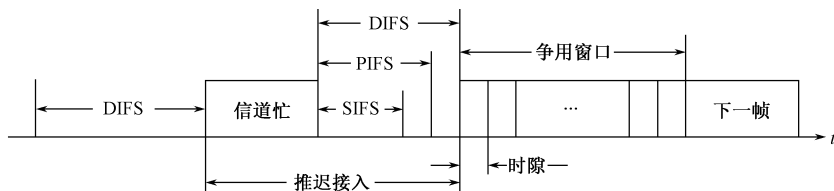


图 11-20 802.11 基本接入方法

IEEE 802.11 允许 DCF 和 PCF 共存,这两种不同机制共存的方法则是通过超帧来实现的,超帧是一个逻辑概念而非实际的帧,它代表一段时间内媒体上的业务量。超帧由一个无争用阶段和一个争用阶段组成,PCF 在无争用阶段控制 BSS 内帧的传输,DCF 在争用阶段控制 BSS 内帧的传输,两者交替控制媒体。图 11-21 表示 PCF 超帧的结构。

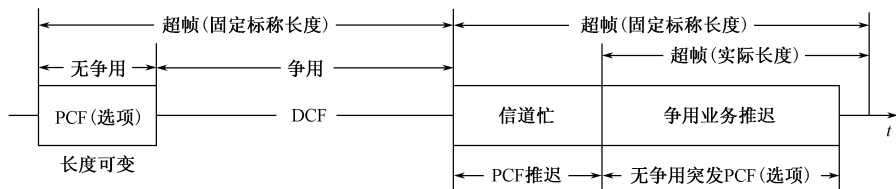


图 11-21 PCF 超帧的结构

假定时间敏感的业务首先占用了信道，轮询由点协调程序进行集中控制，被轮询的站在应答时可使用 **SIFS**，点协调程序收到应答后，继续使用 **PIFS** 询问下一个站，这样就可以一直占用信道。为了防止无限制地进行轮询，无争用阶段的长度必须是受限的，以便留一段时间给后面的争用阶段。假定经过了一段时间信道就被占用，到了下一个超帧的开始，点协调程序就用 **PIFS** 争用信道；若信道空闲，则点协调程序立即接入信道。但信道也可能如图中那样是忙的。这时点协调程序就等待，直到信道空闲时才能接入信道。在这种情况下，超帧的实际长度就缩短了。

## 11.5 无线宽带数据通信协议

### 11.5.1 IEEE 802.16/WiMAX 标准化进程及发展

随着通信技术和新业务的部署，市场与技术的相互作用，传统宽带固定接入用户已经不再满足于仅仅在家庭和办公室等固定环境内使用宽带业务，而是希望使用宽带接入移动服务；另一方面，传统的移动用户也不满足于简单的语音、短信和低速数据业务，希望能使用更高数据速率的业务。用户需求的变化使固定宽带接入服务和移动服务在技术和业务上呈现融合的趋势，宽带移动化和移动宽带化逐渐成为未来通信领域发展的趋势。在移动宽带化方面，3GPP/3GPP2 已经制定了 HSDPA/HSUPA、cdma 2000 1X EV-DO 等技术标准，在移动环境下实现宽带数据传输。在宽带移动化方面，IEEE 802 工作组先后制定了 WLAN 和 WMAN 等技术规范，意图是沿着固定—游牧/便携—移动这样的演进路线逐步实现宽带移动化，其中 IEEE

802.16e/移动 WiMAX 是宽带移动化的重要里程碑，促进了移动宽带化的演进和发展。

IEEE 802.16 标准的提出就是要建立一个全球统一的宽带高速无线数据接入标准。

为了促进标准的发展完善和市场推广，仿照 Wi-Fi 联盟在全球市场成功的模式，2001 年，由世界知名通信企业联合发起并成立了全球微波接入互操作性（World Interoperability for Microwave Access, WiMAX）论坛。WiMAX 论坛是一个非营利性、非官方的工业联盟贸易组织，其目标是对基于 IEEE 802.16 标准的宽带无线接入产品进行兼容性和互操作性认证，完善 IEEE 802.16 标准，推动 IEEE 802.16 技术的产业化，保证所有通过认证的不同厂商的设备之间的互连互通，从而降低生产成本。目前 WiMAX 论坛有 400 多个成员，包括设备制造商、器件供应商、运营商等，形成了完整的产业链。IEEE 802.16 工作组与 WiMAX 论坛之间有着非常紧密的联系与合作，同时又有着分工的不同，前者是标准的制定者，后者是标准的推动者。到目前为止，正式发布的 IEEE 802.16 空中接口系列标准包括 802.16、802.16a、802.16e、802.16d、802.16e 和 802.16f。

根据是否支持移动特性，IEEE 802.16 空中接口系列标准可以分为固定宽带无线接入空中接口标准 802.16、802.16a、802.16c、802.16d、802.16f 和移动宽带无线接入空中接口标准 802.16e。

由于 802.16d 是对 802.16、802.16a、802.16c 的整合和修订，目前主要提及 802.16d 和 802.16e。IEEE 802.16d（即 802.16-2004）是相对比较成熟并且最具实用性的一个标准版本。802.16d 对在 2~66 GHz 工作频段上的固定宽带无线接入系统的 MAC 层和相应多个物理层进行了详细规定，支持视距和非视距传播，支持点到多点（PMP: Point-to-multipoint）和网状网（Mesh）组网。通过对无线发射塔的高度、天线增益、发射功率等参数的设置，单基站（Base Station, BS）最大传输距离约为 50 km。使用 20MHz 信道带宽、64QAM 调制、最高的信道编码效率下，基站每扇区最大传输速率可达 75Mb/s。

IEEE 802.16e 是在固定宽带无线接入标准 IEEE 802.16d 基础上的修订，是 802.16d 标准的增强版，它支持移动台（Mobile Station, MS）以车载速度移动，从而成为新一代 WiMAX 宽带无线接入标准，一般也称为 Mobile WiMAX。作为最新出现的标准，802.16e 采用了很多先进技术，包括：OFDMA、智能天线技术、自适应波束成型、时空码 STC 以及多重输入输出 MIMO 等，并应用于低于 6 GHz 的许可频段。

目前，IEEE 积极推动新一代 WiMAX 标准 802.16j、802.16m，希望借此提升 WiMAX 网络传输效率与传输速率，并开始与主导 3G 标准的国际电信联盟(ITU)沟通合作，希望促成建立 WiMAX 与 3G 互通标准。802.16j 主要为规范移动中继传输设备标准，可解决在既有 WiMAX 网络中出现信号传输死角的问题。而 802.16m 将使移动式 WiMAX 网络传输速率由目前的 100 Mb/s 提升至 1 Gb/s 以上，以此作为让 WiMAX 与 4G 移动通信融合的关键。

### 11.5.2 IEEE 802.16e 协议结构

IEEE 802.16e 标准定义了宽带无线接入系统的空中接口，由媒体接入控制层（MAC）和物理层（PHY）组成。图 11-22 给出了 IEEE 802.16e 协议栈参考模型。其中，SAP（Service Access Point）为服务接入点。服务数据单元（Service Data Unit, SDU）是在两个相邻协议层之间交互的数据单元；协议数据单元（Protocol Data Unit, PDU）是在相同协议层对等实体之间交互的数据单元。



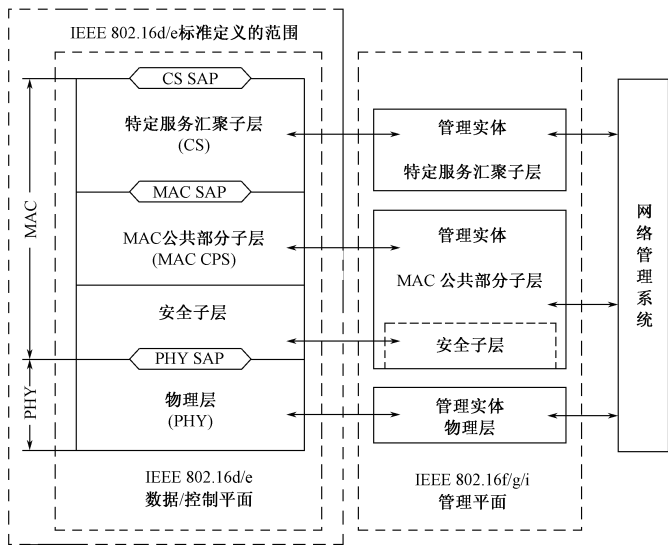


图 11-22 IEEE 802.16 协议栈参考模型

## 1. 物理层

在 802.16e 标准中定义了 4 种物理层实现方式：单载波（Single Carrier，SC）、增强单载波（SCa）、正交频分复用（Orthogonal Frequency Division Multiplexing，OFDM）、正交频分多址接入（Orthogonal Frequency Division Multiple Access，OFDMA）。其中，单载波（SC）调制主要应用在 10~66 GHz 频段，但为了兼容应用于 11 GHz 以下的许可频段的非视距传输，才特别规定了增强单载波（SCa）。

OFDM 和 OFDMA 是 802.16 中最典型的物理层方式。OFDM、OFDMA 方式具有较高的频谱利用率，它使得 802.16 系统在同样的载波带宽下可以提供更高的传输速率。同时，OFDM/OFDMA 方式在抵抗多径效应、频率选择性衰落上也具有明显优势，从而保证 MS 在移动环境中的正常使用。

不同物理层模式分配的信道资源有不同含义，在 SC 和 SCa 模式中，分配的信道资源单元主要指时隙（Time Slot），而在 OFDM 和 OFDMA 中分配的信道资源单元是频域、时域二维资源，包含子载波/子信道和时隙。如果采用了自适应天线系统（Adaptive Antenna System，AAS），资源单元还包含空间资源。

表 11-5 各物理层技术的基本特点

| 物理层类型            | 使用频段          | 基本特点  |
|------------------|---------------|---|
| Wireless MAN-SC  | 10~66 GHz     | 单载波调制，视距传输；<br>可选信道带宽 20 MHz、25 MHz 或 28 MHz；<br>上行采用 TDMA 方式；<br>双工方式可采用 FDD 和 TDD                             |
| Wireless MAN-SCa | 11 GHz 以下许可频段 | 单载波调制，非视距传输；<br>信道带宽不小于 1.25 MHz；<br>上行采用 TDMA 方式，可选支持自适应天线系统（AAS）、自动请求重发（ARQ）和空时编码（STC）等；<br>双工方式可采用 FDD 和 TDD |

| 物理层类型              | 使用频段           | 基本特点  |
|--------------------|----------------|---|
| Wireless MAN-OFDM  | 11 GHz 以下许可频段  | 采用 256 个子载波的 OFDM 调制方式, 非视距传输;<br>可选支持 AAS、ARQ、STC 和 Mesh 等;<br>双工方式可采用 FDD 和 TDD   |
| Wireless MAN-OFDMA | 11 GHz 以下许可频段  | 802.16e 采用 128/512/1 024/2 048 个子载波的 OFDM 调制方式, 非视距传输;<br>信道带宽不小于 1 MHz;<br>可选支持 AAS、ARQ、HARQ 和 STC 等;<br>双工方式可采用 FDD 和 TDD |
| Wireless HUMAN     | 11 GHz 以下免许可频段 | 可采用 SCa、OFDM 或 OFDMA 调制方式;<br>必须支持动态频率选择 (Dynamic Frequencyse Lection, DFS);<br>可选支持 AAS、ARQ、Mesh 和 STC 等;<br>双工方式采用 TDD    |

2. MAC 层

MAC 层功能独立于具体的物理层规范之上, 又可划分为 3 个子层, 由上到下依次为: 特定服务汇聚子层、MAC 公共部分子层、安全子层。

1) 特定服务汇聚子层 (Service-Specific Convergence Sublayer, CS)

CS 子层提供与更高层的接口, 将外部网络数据转换/映射为 802.16 系统内的 MAC 服务数据单元 SDU, 以适配各种上层业务与协议。

IEEE 802.16 MAC 是面向连接的 (Connection-oriented), 汇聚子层将上层业务映射成连接 (Connection)。协议定义了两种汇聚子层: ATM 汇聚子层和数据包 (Packet) 汇聚子层。前者提供对 ATM 业务的支持, 后者提供对基于数据包的业务映射。因为未来移动通信系统将是全 IP 的网络系统。

数据包汇聚子层的核心内容是业务分类。业务分类器 (Classifier) 是一系列映射标准 (Matching Criteria) 的集合, 每个进入 IEEE 802.16 网络的数据包根据分类器定义的规则映射成为连接。MAC 层的每个连接由长度为 16 bit 的连接标识符 (Connection Identifier, CID) 唯一标识。如果一个数据包与某个特定的映射标准相匹配, 那么该数据包将被发送到 MAC 服务访问点 (SAP), 由 CID 所对应的连接进行传输, 对应于该连接的服务流 (Service Flow) 特性对数据包的传输提供了服务质量 (Quality of Service, QoS) 的支持。这种基于连接的机制是提供 QoS 保障的基础。

2) MAC 公共部分子层 (Common Part Sublayer, CPS)

CPS 子层负责执行 MAC 层的核心功能, 包括系统接入、带宽分配、连接建立、连接维护等。

3) 安全子层 (Security Sublayer)

安全子层提供 BS 与 SS (Subscriber Stlion, 用户站) 之间加密、鉴权、密钥交换等与安全有关的功能。

### 3. MAC 层对物理层的支持

#### 1) 双工方式

IEEE 802.16 MAC 层支持频分双工 (FDD) 和时分双工 (TDD) 两种双工方式。双工方式的选择会影响物理层的一些参数, 进而影响与这些参数相关的物理层特性。

在频分双工系统中, 上、下行链路分配不同的工作频率。上行和下行链路中每一帧的持续时间固定, 这有利于使用不同的调制方式, 也简化了带宽分配算法。

在时分双工系统中, 上行和下行链路共享同样的频率, 在不同的时间发送数据。每一帧有固定的时间长度, 分为下行子帧 (Downlink Subframe) 和上行子帧 (Uplink Subframe), 下行子帧在先, 上行子帧在后。一个帧被分成整数个物理时隙 (Physical Slot, PS), 这有助于带宽的划分。上行链路和下行链路之间的分界是一个系统参数, 这个参数由系统上层来控制的。每一帧中上、下行子帧占用带宽的比例可以自适应调整。在 PMP 模式中, 资源的调度和分配由 BS 集中控制, 能够根据上下行数据量灵活动态地分配带宽, 对于上下行不对称业务具有较高的资源利用率。在下行和上行子帧之间插入 BS 发送/接收转换间隔 (Transmit/Receive Transition Gap, TTG), 而在上行子帧和下一个帧之间插入 BS 接收/发送转换间隔 (Receive/Transmit Transition Gap, RTG), 以留出必要的保护时间。

#### (2) DL-MAP/UL-MAP 管理消息和物理成帧。

IEEE 802.16 系统使用突发 (Burst) 数据传输模式。在形成物理层帧的过程中, BS 的 MAC 层需要产生下行链路映射 (DL-MAP) 管理消息和上行链路映射 (UL-MAP) 管理消息, 为 SS 访问下行子帧和上行子帧提供相关信息。在 PMP 模式的系统中, DL-MAP 和 UL-MAP 消息由 BS 以广播方式发送给其覆盖范围中的所有 SS。UL-MAP 消息紧接在 DL-MAP 消息之后。

DL-MAP 消息定义了基于突发的物理层中下行传输间隔的使用, 指明下行链路带宽资源的分配。也就是说, DL-MAP 消息定义的是 SS 如何来接收下行链路上的信息。PMP 网络中, 下行子帧中的信息被广播发送给 BS 覆盖区域内的所有 SS。DL-MAP 管理消息并没有显式地指出下行子帧中的各部分数据都分别属于哪个确定的 SS。每个 SS 接收到信息后, 通过提取、检查接收到的 MAC PDU 中的连接标识符 (CID) 能够确定发往该 SS 的数据, 丢弃不属于自己的数据。

UL-MAP 消息根据突发相对于分配开始时间 (Allocation Start Time) 的偏移定义了上行传输间隔的使用, 指明上行链路的带宽资源分配。上行子帧中包括用于初始化测距的间隔、竞争带宽请求的间隔、上行数据传输的间隔等等, 这些间隔的具体分配都是由 BS 的 MAC 层统一控制调度分配, 并在 UL-MAP 消息中指出。

BS 还需要周期性广播下行信道描述符 (Downlink Channel Descriptor, DCD) 管理消息和上行信道描述符 (Uplink Channel Descriptor, UCD) 管理消息, DCD 消息指明应用于当前 DL-MAP 消息的下行突发属性 (Downlink Burst Profile); 而 UCD 消息指明应用于当前 UL-MAP 消息的上行突发属性 (Uplink Burst Profile)。DL-MAP、UL-MAP、DCD 和 UCD 消息都位于帧控制头部分。通过 MAC 层的协调, BS 和 SS 可以根据需要灵活地改变每一帧的突发类型, 从而选取适当的物理层传输参数 (调制方式、编码方式、发射功率等)。

IEEE 802.16e 系统重点关注 OFDM/OFDMA 物理层技术。OFDMA 的子信道化可以使系统扩大覆盖、提高容量。802.16e OFDMA 标准中针对不同的信道带宽定义了不同的子载波数

目，从而可以使系统在抵抗频率选择性衰落和支持移动性方面取得一个很好的折中。

## 思考与练习

- 11-1 解释通信网络协议组成要素。
- 11-2 试画出 OSI 参考模型，并简述各层功能。
- 11-3 试说明两个系统的应用进程通信时，各层数据的传送过程。
- 11-4 物理层协议中规定的物理接口的基本特性有哪些？并说明其基本概念。
- 11-5 什么是数据链路？数据链路传输控制规程的功能有哪些？
- 11-6 数据链路的结构有哪几种？
- 11-7 画出面向字符型的传输控制规程中信息报文（信息文电）的基本格式。
- 11-8 在面向字符型的传输控制规程中说明轮询/选择方式的通信过程。
- 11-9 在面向字符型的传输控制规程中为了能透明传输，采取什么措施？
- 11-10 画出 HDLC 的帧结构，并说出各字段的含义。
- 11-11 HDLC 中对一帧中的哪些字段要进行“0”比特插入后传送？若已知 I 字段的内容为 0111111111011110，试给出“0”比特插入后的比特序列。
- 11-12 HDLC 规程的 3 种操作方式的特点是什么？
- 11-13 HDLC 规程中数据链路结构有哪几种？
- 11-14 HDLC 传输过程，试给出 NRM 方式双向同时传输数据链路的数据传输过程。
- 11-15 HDLC 传输过程，试给出一差错的恢复过程。
- 11-16 画图说明无线信道的隐终端和暴露终端问题。
- 11-17 画图说明载波侦听多址访问/碰撞回避机制。
- 11-18 画图说明 IEEE 802.11 基本接入方法。

## 第 12 章 无线数据通信组网技术介绍

在过去的几十年期间,无线数据通信按照自己的规则蓬勃发展,从蜂窝电话网到无线接入 Internet 和无线家庭网络等,无线数据通信给我们的生活带来了深刻影响,在经过呈指数增长后,今天的无线数据通信产业已经成为世界上最大的产业之一。

### 12.1 无线数据组网的发展

无线数据通信组网主要包括网络体系结构的设计和网络协议的设计,前者需要解决网络拓扑结构、网络控制管理、协议体系结构顶层设计问题,后者主要针对所设计的网络体系结构进行有效拓扑控制算法、管理协议、通信协议等底层技术的实现。算法协议的性能表现是与其应用背景及条件密切相关的,在一定条件下表现优良的算法和协议在条件改变之后可能变得并不适用,因此,算法和协议优化需要根据应用背景来确定,没有普遍适应的网络算法和协议。

无线通信(系统)网络的应用从逻辑上可以分成两类:面向语音的应用和面向数据的应用。每一类都有局域和广域两个应用场合,从而形成 4 个不同的应用部分。面向语音的无线网络应用围绕着连接 PSTN 的无线网络,这些业务进一步形成局域网应用和广域网应用。面向数据的无线网络应用围绕着 Internet 和计算机通信网络的基础结构。这些业务进一步分为宽带局域与 Ad Hoc(点对点)应用和广域无线数据应用,其发展过程已经有 4 代。

第一代(1G)无线通信系统,是面向语音的模拟无线系统,使用 FDMA 技术实现,能提供基站和移动用户间的模拟话音和低速率数据通信。典型的标准有美国的 AMPS、欧洲的 TACS。

第二代(2G)无线通信系统,是面向语音的数字无线系统,使用 TDMA 或窄带 CDMA 技术实现。与第一代网络相比,第二代网络增加了用来传输寻呼与其他数据业务的功能,如传真、较高速数据接入等。网络控制功能则分散于网络中,移动站承担了更多的控制功能。网络中的移动单元有许多第一代网中用户单元没有的功能,如接收功率报告、邻近基站搜索、数据编码以及加密等。典型的语音业务标准有:欧洲的 GSM、CT-2 和 DECT,北美的 IS-54 (IS-136)、IS-95 和 PACS,日本的 JDC、PHS 等。典型的数据业务标准有 CDPD、GPRS、EDGE、IEEE 802.11、HiperLAN 等。

第三代(3G)无线通信系统,是面向宽带多媒体无线网络系统,可实现高速语音、图像、数据及无线互联网等业务。ITU 规定的第三代移动通信系统无线传输技术必须满足以下 3 种环境的最低要求,即高速移动环境,最高速率达 144 kb/s;室内外低速移动环境,最高速率达 384 kb/s;室内固定或低速步行环境,最高速率达 2 Mb/s。典型的 3G 标准:WCDMA、CDMA 2000 和 TD-SCDMA 以及 WiMAX。

第四代(4G)无线通信系统,是目前在研的下一代无线通信系统,国际电信联盟(ITU)已提出 4G 系统的目标,预期将具有以下特点:

(1) 全世界通用标准的系统,且可在现有不同的无线通信系统下运作。

(2) 以 IP (Internet Protocol) 为主,用户在任何时间、任何地点都能使用 4G。在高速移

动下，需达到 100 Mb/s 的传输速率。而在慢速状态下，传输速率需达到 1 Gb/s。

(3) 4G 系统不但支持固定的无线传输，亦支持移动的无线传输且依实际需要可在固定与移动网络之间互相切换。

(4) 4G 不但可以解决第三代无线通信 3G 的缺点，并且能提供更多元化的无线宽带服务，例如，更逼真的语音，更高清的影像及更快的下载速度。用户容易连上网络，并可依照个人的喜爱选择所需的服务。

(5) 4G 所提供的网络服务与相关设备的价格为一般用户可接受。

## 12.2 无线数据网络体系结构

### 1. 网络拓扑构造

无线数据网络拓扑结构是指为构成无线通信系统组织结构而建立的系统节点的布局及其相互间的结构方式（连接关系）。无线网络拓扑结构规划通常需要考虑网络规模、业务强度及类型、网络控制管理效能、对节点间连接关系动态变化的适应能力等。目前，比较典型的无线网络结构主要有平面结构和分级结构。

在平面结构无线网络中，各节点的功能和地位平等，不需要复杂的网络结构维护过程。同时，由于网络中各节点地位对等，使得网络中原则上不存在瓶颈而比较健壮，对网络使用环境的适应能力也比较强。但平面结构网络中，当网络规模加大时维护这些动态变化的路由需要大量的控制消息，路由维护的开销呈指数增长而消耗掉有限的带宽。随着用户增多，移动性加强，平面结构网络需要维护的控制开销快速增加，网络性能急剧下降。因此，平面结构网络的可扩展性较差，主要适用于网络规模较小的场合。

在分级结构中，无线网络通常被划分为若干群，每个群由一个群首和多个群成员组成。群首节点负责群内节点管理，实现群间数据接发。群首形成高一级的网络拓扑，依次可以形成多级拓扑结构。在分级结构中，群首可以预先指定，也可以由节点使用分群算法自动选举产生，并根据节点连接关系的变化重新分群以构建网络拓扑。在分级结构网络中，群成员不需要维护复杂的路由信息，大大减少了网络中路由控制消息的数量，网络具有很好的可扩展性，路由算法设计也较为灵活，可采用先验式和反应式路由协议相结合的方式提高路由算法性能。

综上所述，当无线网络规模较小时，可采用平面结构进行组网；当无线网络规模较大时，可采用分级网络结构进行组网。

### 2. 网络协议体系构造

目前，在计算机数据通信网领域内，最具权威的标准是国际标准化组织（ISO）提出的开放式系统互连参考模型（OSI/RM）。OSI/RM 把计算机网络在功能上分为 7 个层次，每一层都执行特定的功能，相邻的上下两层之间通过层间的服务访问点（软件接口）进行通信，上一层利用下一层功能提供的服务。当前，无线数据网络构建过程中所共同采用的网络协议模型，就是开放系统互连的七层参考模型。

无线数据网络的协议体系结构设计，在遵循开放式系统互连协议模型的同时，还要结合无线数据网络应用背景情况下的系统服务需求进行改进，并与网络拓扑结构相适应，应包含

多个协议栈，涵盖各级无线数据网络内部组网协议栈、各级网络互联协议栈、多网网关协议栈等，从而实现整个无线数据网络建立、网络维护和网络管理功能，满足无线数据使用要求。

## 12.3 蜂窝网络拓扑

蜂窝网络拓扑结构是一种特殊的多基站（BS）网络基础结构配置，使用了频率复用概念。目前，蜂窝网络拓扑是在大规模陆地无线网络和卫星无线网络中使用的主要拓扑形式，蜂窝通信的概念最早由贝尔实验室在 20 世纪 70 年代提出，它使有效的带宽可以容纳巨大数量的用户。

### 1. 蜂窝概念

蜂窝无线通信是指在传输中使用大量低功率的基站（BS），每个基站（BS）只覆盖有限的区域。用这种方式，每次建立一个新的基站（BS）时，容量就会增加，因为同样的频谱在指定的区域内可以被复用几次。蜂窝的基本原理是把覆盖区域分为大量相连的小区域，每个小区域都是使用自己的无线基站。对这些小区域以智能的方式分配信道，这样可以减小干扰、提供充分的性能以及满足这些区域的通信量。这种小区域被称为小区。一组小区组成区群。在每个区群中使用相同的频谱，因此可以把整个频带分成频带组，并分配给区群中的小区使用。在区域中无线频率组的空间分布必须采用某种方式，以便获得理想的性能。这是蜂窝无线通信网络规划的重要内容，一个区群中小区的数量称为区群大小或频率复用因子。图 12-1 为蜂窝网络拓扑结构示意图。

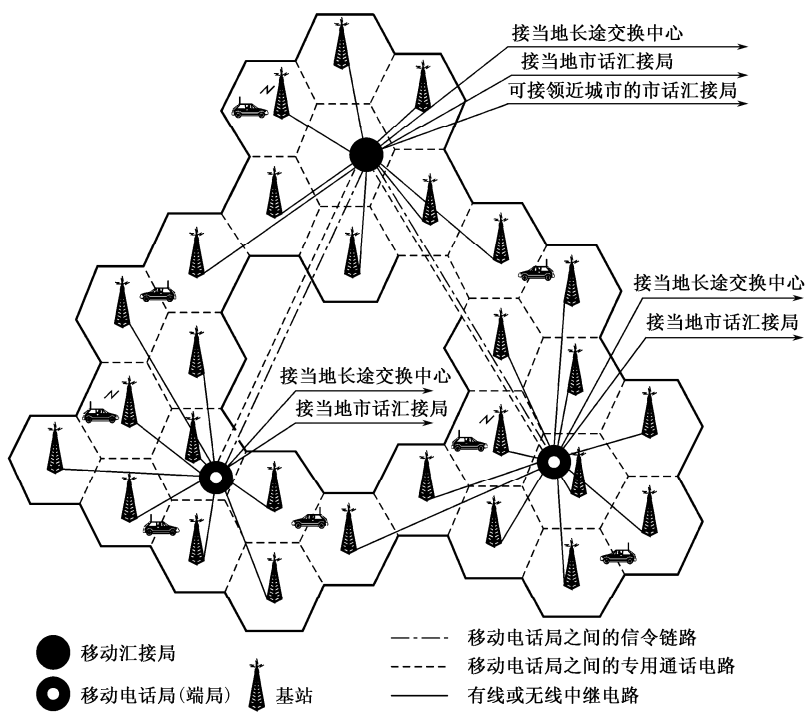


图 12-1 蜂窝网络拓扑结构示意图

蜂窝结构拓扑能有效地增加可用频率所能支持的用户数量。例如在一个城市中，可供使用的总带宽为 25 MHz，每个用户需要 30 KHz 带宽。如果使用单个天线覆盖整个城市，就只能同时支持  $25\text{ MHz}/30\text{ kHz}=833$  个用户。现在使用蜂窝结构，合理分布 20 个低功率天线，以减少各种干扰。将可用频带分成 5 部分，每个小区分配一部分，每个小区分配到的频谱为  $25\text{ MHz}/5=5\text{ MHz}$ 。使用包含 5 个小区的区群，每个小区同时支持的用户数量为  $5\text{ MHz}/30\text{ kHz}=166$ ，每个区群的用户数量为  $5\times 166=830$ 。如果城市中有 4 个包含 5 个小区的区群，则同时可支持的总的用户数量为  $830\times 4=3\,320$ ，约为单个天线所支持容量的 4 倍。

## 2. 蜂窝小区的分类

在现代蜂窝网络中，出现了一系列支持各种不同应用的小区，这些小区定义如下：

(1) 大小区：用于覆盖全国性的区域，覆盖范围为几百公里，主要应用在卫星通信方面。

(2) 宏小区：用于覆盖大城市地区，覆盖范围为几公里，天线安装在覆盖区主要建筑物的房顶上。

(3) 微小区：用于覆盖街道之内，覆盖范围为几百米，天线高度低于沿街两侧建筑物房顶的高度。

(4) 微微小区：主要是在建筑内使用的小型小区，支持本地室内网络，覆盖范围为几十米内，如无线局域网 (WLAN)。

(5) 毫微微小区：这是蜂窝中最小的单元，用于连接个人设备，覆盖范围只为几米以内，如笔记本电脑和蜂窝电话。

## 3. 容量扩展技术

在 20 世纪 90 年代，蜂窝移动电话发展非常迅速，为适应激烈的市场竞争和获取更大的运营利润，产生了不少方法用于扩展蜂窝移动电话网络。目前，用于容量扩展的方法有 3 种。

第一种方法是改变蜂窝结构。结构上的方法包括小区分裂、小区划分扇区使用定向天线、微小区区域技术和使用多个复用因子。它们通过增加小区站点或修改天线特性改变了小区覆盖的大小和形状，从而增加了能量。这些技术不需要额外的频谱，也不需要无线 Modem 或系统接入技术作任何重大改变，因而不需用户购买新的终端。这些特点证明了它们是一种实用的、廉价的扩展网络容量的方法。

第二种方法是改变频率的分配方案，根据不同小区流量需求动态地分配频率。每个小区的通信量都随着服务区的地形和通信量有关的时间动态地变化。例如，在大部分城市地区，高峰时间的通信量最大，而晚上或节假日的通信量相对较小；而在住宅地区恰与此相反。这样将信道动态地分配给不同的小区，就可增加网络的容量。这些技术不需要改变终端或系统的物理体系结构，只需要改变相应的软件程序就可以。

第三种方法也是最有效的扩展网络容量的方法就是改变 Modem 和接入技术。蜂窝移动产业的空中接口最初使用 FM 模拟调制技术，现在发展成为使用 TDMA 和 CDMA 技术，即可增加网络容量，同时也为语音和数据业务的结合提供了一个良好的环境。但是，这种技术上的改变既需要用户购买新的终端，又需要服务提供商安装新的基础设备。

# 12.4 移动 Ad Hoc 网络技术

“Ad Hoc”一词源自拉丁语，原意为“for the specific purpose only”，即“特别的、临时



的”。移动 Ad Hoc 网络就是移动自组网 (Mobile Ad Hoc Network, MANET)，由一组无线移动节点组成的、能快速部署的、自组织的临时性分布式网络。

移动自组网 (MANET) 的前身是分组无线网 (Packet Radio Network, PRNET)，到目前为止，已经有 30 多年的发展历史。早在 1972 年，源于对军事通信的需要，美国国防部高级研究计划署 (DARPA) 就启动了 PRNET 项目，研究目标是将数据分组交换技术引入到无线环境中，开发军用无线数据分组网络。PRNET 采用分布式体系结构，采用主动多跳路由技术，支持 ALOHA 与 CSMA 两种 MAC 协议，支持动态共享广播无线信道。1983 年，DARPA 又启动了抗毁性自适应网络 (SURvivable Adaptive Network, SURAN) 项目，主要是将 PRNET 成果加以扩展，以支持更大规模的网络。1994 年，为了使全球信息基础设施支持无线移动环境，DARPA 启动了全球移动信息系统 (Global Mobile Information System, GloMo) 计划。目标是支持无线节点之间随时随地的多媒体连接，解决移动 Ad Hoc 网络的 3M 问题：移动 (Mobile)、多跳 (Multihop) 以及多媒体 (Multimedia)。1997 年 6 月，Internet 工程任务组成立了 MANET (Mobile Ad Hoc Network) 工作组，主要致力于移动 Ad Hoc 网络路由协议的标准化工作，极大地推进了商用移动 Ad Hoc 网络的研究与开发。

移动 Ad Hoc 网络不需要固定基站支持，实现分布式的无中心管理，可临时组织，具有高度移动性，网络抗毁与快速部署能力强，适合野战通信等特殊环境使用要求。

## 1. MANET 特点

与其他类型的无线网络 (如蜂窝网、卫星网) 相比，MANET 网络具有下列显著特点：

(1) 全分布式。MANET 网络完全由无线节点组成，没有任何固定基础设施。网络的运行、组织和管理都是分布式的，且可以自组织。

(2) 动态拓扑。节点可以随机移动，网络拓扑在任意时刻可能发生不可预知的变化。

(3) 多跳拓扑。无线节点通信距离有限，当目标节点超出源节点通信范围时，需要中间节点进行中继转发，这就构成了多跳拓扑。

(4) 带宽有限且易变。无线链路的用户带宽有限，尤其在野战环境下，由于复杂地形引起的多径、衰落、噪声以及敌方干扰的影响，无线链路容量容易发生变化。

(5) 能源有限。无线节点主要依靠电池或其他可耗尽能源来供电，这使得节点的能源有限，也限制了节点的通信和计算能力。

总之，MANET 网络是一种移动、无线、多跳的分布式网络。这些特点使得开展 MANET 网络组网技术研究面临许多挑战。

## 2. MANET 网络的研究内容

### 1) 媒体访问控制 (MAC) 技术

MAC 技术是无线网络组网协议的基础，它用于解决各用户无冲突地访问无线信道、传输数据。对于共享广播信道的 MANET 网络而言，MAC 协议对网络整体性能起着决定性作用。由于 QoS 保障要求、分布式多跳环境等原因，以及隐式终端和暴露终端的存在，它的设计复杂性也远超过一般的集中式单跳无线网络。在评估 MANET 网络的 MAC 协议时，一般需要考虑的性能指标有：吞吐量、时延、公平性、能量效率以及服务质量 (Quality of Service, QoS) 支持等。

## 2) 路由协议

路由协议是 MANET 网络的核心问题，也是目前研究最广泛的内容。由于节点移动、动态拓扑、无线链路带宽窄且易变等原因，有线网络的路由协议难以满足 MANET 网络要求。目前，针对 MANET 网络路由协议已经开展了大量研究工作，并取得了不少成果。评估 MANET 网络路由协议的性能指标一般包括：端到端的数据吞吐量与时延、路由建立时间、协议开销、单向链路支持、收敛速度以及 QoS 支持等。

## 3) 分群算法

网络分群是分级网络结构需要重点考虑的问题，网络分群协议包括移动分群算法和分群保持策略。分群算法的性能评价主要在一定网络节点规模和节点无线覆盖距离情况下，通过分群数目、网关数目、节点到节点之间的跳数、连通性、群的平均大小、平均邻居节点数等参数进行综合考察。

## 4) QoS 保障问题

随着应用的不断扩展，MANET 网络需要支持语音、图像等多媒体业务，从而对带宽、时延、时延抖动等 QoS 保障提出了很高要求。MANET 网络不可预测的无线链路特性以及节点移动性，都为其 QoS 保障带来了比有线网络更复杂的挑战，使得 QoS 问题成为 MANET 网络的一个研究热点。MANET 网络的 QoS 问题是一个系统性问题，需要各层都提供相应的保障机制。如 MAC 层要提供资源预留，网络层要提供 QoS 路由，应用层要提供自适应编码与压缩技术等。

# 12.5 无线 Mesh 网络技术

无线 Mesh 网（Wireless Mesh Network，WMN）又称为无线网状网和无线网格网，是基于 IP 协议的大容量、高速率、覆盖范围广的无线网络，无线 Mesh 网络是在移动自组织网络（MANET）和无线局域网（WLAN）基础上发展起来的一项网络技术，通过呈网状分布的无线接入点（AP，AccessPoint）间的相互合作和协同，成为宽带接入的一种有效手段，它作为下一代因特网核心网的无线版本，有效地解决“最后一公里”的瓶颈问题，是一种具有动态自组织、自配置、高速率、高容量等特性的分布式宽带无线网络。

无线 Mesh 网更主要的是一种网络架构思想，主要功能体现在无中心、自组网、多跳连接和路由判断选择等，具有与现有无线网络的兼容性及互操作性。

## 1. Mesh 网络的组成

无线 Mesh 网络（WMN）由用户节点（终端）、无线 Mesh 路由器（Wireless Router，WR）节点和网关节点组成。但 Mesh 终端既可以是工作站也可以是移动终端设备，根据网络具体配置的不同，WMN 不一定含以上所有类型的节点。

无线 Mesh 网络可以分为骨干结构 Mesh 网络、终端结构 Mesh 网络和混合结构 Mesh 网络。

在骨干结构 Mesh 网络中，Mesh 路由器与终端之间的连接一般通过无线方式，而各 Mesh

路由器之间的连接既可以采取无线方式也可以采用有线方式。Mesh 路由器与终端间、Mesh 路由器之间可以采用多种传输技术和频段进行传输数据，目前以 IEEE 802.11 技术居多。

终端结构 Mesh 网络的网络结构比较类似于 Ad Hoc 无线自组织网的网络结构，各节点是对等的自组织形式。各节点间存在端对端（Peer-to-Peer）网络链路。在该结构下，可以没有专用的 Mesh 路由器。终端型 Mesh 网络一般使用同频段所示。

混合结构 Mesh 网络综合了以上 2 种结构的优点，是 Mesh 网络中最灵活、最方便的形式，兼容性和扩展性强，其网络结构如图 12-2 所示。

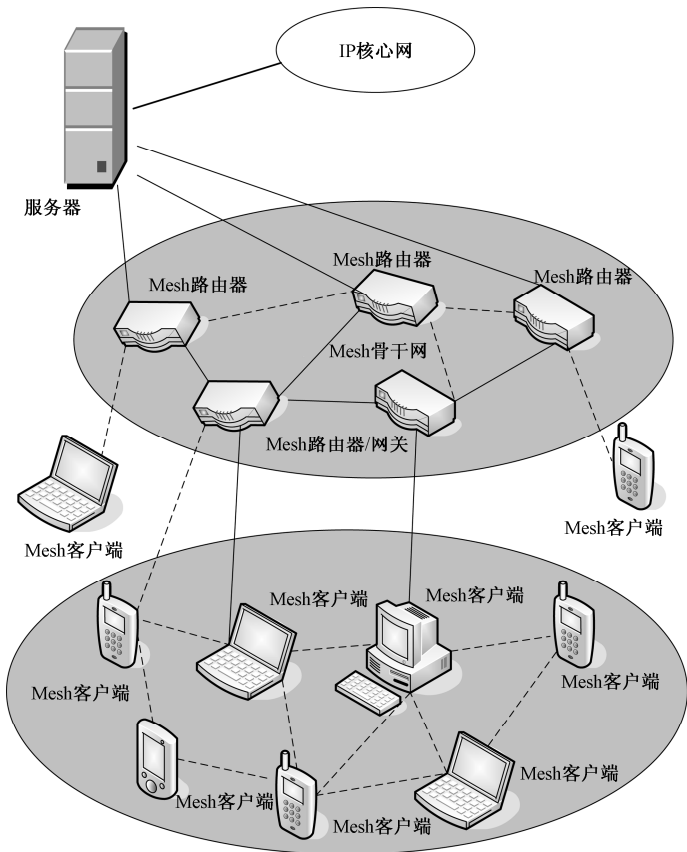


图 12-2 混合结构 Mesh 网络示意图

## 2. 无线 Mesh 网络的特点

无线 Mesh 网络（WMN）是多跳与多点到多点结构的融合，具有以下几个重要特点：

（1）多跳的结构。在不牺牲信道容量的情况下，扩展当前无线网络的覆盖范围是 WMN 的最重要的目标之一。WMN 的另一个目标是为处于非视距范围的用户提供非视距连接。Mesh 网络中的链路比较短，所受干扰较小，因此可以提供较高的吞吐量和较高的频谱复用效率。

（2）支持 Ad Hoc 组网方式，具备自形成、自愈和自组织能力。WMN 灵活的网络结构、便利的网络配置、较好的容错能力和网络连通性，使得 WMN 大大提升了现有网络的性能。在较低的前期投资下，WMN 可以根据需要逐步扩展。

(3)移动特性随 Mesh 节点类型的不同而不同。Mesh 路由器通常具有较小范围的移动性，而 Mesh 客户端既可以是静止不动的节点，也可以是移动的节点。

(4)支持多种网络接入方式。WMN 既支持通过骨干网接入的方式，又支持端到端的通信方式。此外，WMN 可与其他网络集成，为这些网络的终端用户提供服务。

(5)对功耗的限制取决于 Mesh 节点的类型。Mesh 路由器通常没有严格的功耗限制，但 Mesh 客户端需要有效的节能机制。

(6)与现有无线网络兼容，并支持与 WiMAX、Wi-Fi 和蜂窝网络等的互操作。  
这种使得 WMN 具有以下优点：

- 自配置：节点之间通过开放的无线链路，形成单跳或多跳连接，自动完成组网；
- 自调节：节点之间拥有多条通信路径，业务可以灵活地选择合适的（如最短路径、最少干扰、最快速率等）路径进行传输；
- 自愈：当某一节点出现故障受损或一条链路出现拥塞，网络中的业务可以选择绕开相应的节点或链路，网络的可靠性增强。
- 可扩展性：可以方便地添加或删除网络节点，调整网络覆盖范围，降低系统的建设和管理成本。

### 3. 无线 Mesh 网络的应用

无线 Mesh 网络已有一定程度的商业应用，目前的应用主要为 Internet 接入和内部端到端通信两类，包括：宽带家庭应用、社区网络互连、企业接入、城域网络互连。目前无线 Mesh 网络已遍布了无线市政、智能交通、制造业、能源行业、军队系统等几乎全部重要行业。

### 4. 无线 Mesh 网络关键技术

WMN 设计中的一个关键问题是，开发能够在两个节点之间提供高质量、高效率通信的路由协议。由于其网络节点的移动性使得网络拓扑结构不断变化，传统的 Internet 路由协议无法适应这些特性，需要有专门的、应用于无线 Mesh 网络的路由协议。在路由协议设计时，要考虑以下几方面。

选择合理的路径算法。现有的很多路由协议是以最小跳数为标准选择路径的，但是，如果连接质量较差或者网络拥塞的话，这种标准就不合适了。因此，在选择路径时，不能只考虑最小跳数，还应该综合考虑网络的连接质量和往返时延等因素。

确保对连接失败的可容错性。WMN 的目标之一就是在出现连接失败时，确保网络的健壮性。如果一个连接失败了，路由协议必须很快选出另外一条路径，避免出现服务中断。

实现网络负载平衡。采用 WMN 的另一个目标是实现用户对资源的共享。当 WMN 中的某一部分出现数据拥塞时，新的数据应该选择流量较小的路径。

能够同时满足不同类型节点的需求。对于路由器，它的移动性较弱且没有能源消耗的限制，它所需要路由协议比现有的 Ad Hoc 网络的路由协议要简单得多；对于终端用户来说，情况恰恰相反。在设计 WMN 路由协议时，要充分考虑这两种节点的差异，分别满足它们的不同需求。

如果仅考虑提高某一层面协议的性能，效果并不明显。目前，WMN 发展的趋势是跨层设计，即同时考虑多个层面间的影响。WMN 的跨层设计要求打破传统的 OSI/RM 参考模型

中严格分层的束缚，针对各层相关模块/协议的不同状态和要求，利用层与层之间的相互依赖和影响，对网络性能进行整体优化。具体来说，跨层设计就是充分、合理地利用现有的网络资源，达到系统总吞吐量的最大化、总传输功率的最小化、QoS 的最优化等最终目的。

## 思考与练习

- 12-1 试说明移动 Ad Hoc 网络技术的基本原理。
- 12-2 试说明无线 Mesh 网络的特点，画出混合结构 Mesh 网络示意图。
- 12-3 何谓无线数据网络拓扑结构？试说明平面结构和分级结构的区别。

# 第 13 章 战术数据链系统与技术

## 13.1 战术数据链基本概念

### 13.1.1 何谓战术数据链

“数据链”一词，最早起源于 20 世纪 50 年代末，当时即为北大西洋公约组织（NATO）及美国海军所使用。它主要是采用无线网络通信技术和应用协议，实现机载、陆基、舰载的指挥和控制时的战术数据信息交换，其应用示意图如图 13-1 所示。



图 13-1 战术数字信息链路应用示意图

数据链是链接数字化战场上的指挥中心、作战部队、武器平台的一种信息处理、交换和分发系统。在以计算机为核心的现代自动化军事指挥控制通信与情报系统中，专门用于数据传输的设备和无线电链路的总和，北约称为数据链路（Link），美军称之为战术数字信息链路（Tactical Data Information Link, TADIL），外军中常称为“战术数据链”（Tactical Data Link）。

美国国防部的标准军事术语词典对于战术数据链的定义：战术数据链是由军方标准化并且用于传输数字信息的通信链路，采用一种或多种网络结构及多种传输手段，连接两个以上指控、传感器或武器系统，用于交换战术信息。

我国定义：数据链是以无线传输为主，按照统一的消息标准和通信协议，链接传感器平台、指挥控制平台和武器平台，实时处理和分发战场态势、指挥引导、战术协同、武器控制等格式化信息的系统。

通俗地说，数据链是一种高效传输、实时分发格式化消息的信息链路。

数据链综合利用各种调制解调技术、纠错编码技术、组网通信技术及信息融合技术，来适应指挥控制系统之间数据通信的需求，并利用无线信道在各级指挥所和各作战平台之间构

成一体化的数据通信网络，按照规定的信息格式，实时、自动、保密地传输和交换各种作战信息。

数据链的三大基本要素：格式化消息、通信协议和数据链设备。格式化消息是数据链系统传送的数据内容，它确定了机器可以识别的格式化信息，传输的数据可以用于控制武器平台，也可以产生图形化的人机界面；通信协议包括频率协议、链路协议、网络协议和加密标准等，主要用来建立特殊通信链路，并控制数据链路中的传送；数据链设备主要有战术数据系统终端的消息处理器、加解密设备、网络控制器和传输设备。

例如，战术数据链系统将雷达、侦察卫星等传感器平台收集的信息，或者指挥员、操作人员发出的各种数据，编排成标准的格式化消息；加解密设备负责加密发送信息和解密接收到的信息。网络控制器进行消息调制、加密、检错与纠错等，将格式化消息变成符合通信设备传输要求的数据信号，或者与指挥控制系统、武器控制系统进行信息交换；传输设备传递、交换和处理数据信息。

目前，世界上已有美国、俄罗斯、以色列、北约组织和中国台湾地区的军队均装备有数据链。比较典型的有：美国和北约的 TADIL B (Link 1)、TADIL C (Link 4)、TADIL A(Link 11)、TADIL J(Link 16)及正在改进的可工作多频段的 Link 22 号数据链；俄罗斯的蓝天、蓝宝石，以色列的 S 波段数据链；中国台湾引进美军的 16 号数据链等。其中，美军对数据链的开发和应用最为广泛。几种常见数据链的特性如表 13-1 所示。

表 13-1 几种常见数据链的特性

| 链 路  | 结 构          | 协 议   | 报 文 标 准      | 数据速率/（b/s）                            |                       |                              |
|------|--------------|-------|--------------|---------------------------------------|-----------------------|------------------------------|
|      |              |       |              | 战术信息                                  | 附带<br>奇偶校验            | 附带误差<br>检测与修正                |
| 4A 号 | 时间分隔<br>多路传输 | 指令/响应 | V 序列<br>R 序列 | 3.06                                  | —                     | —                            |
| 11 号 | 网状结构         | 网控呼叫  | M 序列         | 快 1.80<br>慢 1.09                      | —<br>—                | 2.250<br>1.364               |
| 16 号 | 时间分隔<br>多路存取 | 预定时隙  | J 序列         | 标准 26.88<br>组合 2 53.76<br>组合 4 107.52 | 28.8<br>57.6<br>115.2 | 59.520<br>119.040<br>238.080 |

13.1.2 战术数据链功能及特点

数据链的基本作用与功能，是将探测网络、指控网络和武器平台有机的交链在一起，构成海、陆、空、天一体化的信息链路，实现战场态势共享，实时信息交换，从而提高协同作战能力和整体作战效能，具体包括：

- 支持实时信息共享，提高态势感知能力；
- 构建“侦、判、打、评”闭合回路，提高指挥控制能力；
- 实现“传感器到射手短路”作战，提高快速行动能力；
- 区域集成一体化的作战体系，提高联合作战能力。

为了实现以上各项功能，数据链系统通常具有数据链路平台一体化，传输内容格式化、信息传输实时化和时间空间一致化。数据链路平台一体化是将一般通信系统的“人—机—人”

工作方式转变为“机—机”工作方式。传输内容格式化是数据链的灵魂。时间空间一致化即对异步观测数据进行时间和空间对准。信息传输实时化即要求可靠性服从于实时性，这时应压缩信息量，提高信息表达效率，选用传输效率高、简单实用的通信协议，并采用相对固定的网络结构和直达的信息传输路径，在实时性极高的应用场合可直接采用点到点的链路传输，综合考虑实际信道的传输特性、信号波形、通信协议、组网方式和消息标准等技术环节，统一进行设计，从而提高数据传输的速率，缩短各种机动目标信息的更新周期。

数据链路具体特点如下：

(1) 信息传输的实时性。数据链系统传输的信息是直接为指挥和武器控制系统提供支持和服务的。面对瞬息万变的战场形势，如果不能在规定时间内把相应战术信息传递给所需的作战单元，数据链就失去了存在的意义。因此，提到数据链必须强调信息传输的实时性。

(2) 信息传输的可靠性。数据链系统主要通过无线信道来传输信息，而工作在复杂电磁环境下的无线信道是一个变参信道，在信号传播过程中，存在着各种干扰、衰落现象，严重影响信号的正常接收。在语音通信时，收信人员可以借助听觉判断力，从被干扰的信号中正确识别信息。对于数据通信来说，信道质量的下降会导致误码率升高，当不能满足数据的接收门限时，就会无法完成战术数据的传输。为了克服、减轻这些干扰情况，数据链系统采用了高性能的纠错编码技术来降低数据传输的误码率和均衡技术来克服信道衰落现象。

(3) 信息传输的安全性。为了不让敌方截获己方信息，使己方信息安全秘密的传输，数据链系统一般采用严格的保密措施，如加装保密机对数据进行加密处理，采用跳、扩频等抗干扰电台传输，降低敌方的截获概率，确保己方的信息安全。

(4) 信息格式的一致性。由于现代战争的复杂性，在协同作战的通信中，为了实现多兵种、多网络间信息的共享与互通，各作战系统和平台间需要不断交换信息，为避免在不同网系间因格式差异而引起信息误解等问题，数据链系统具有一套相对完备统一的消息标准，使网络中的不同成员对消息的语义有相同的解释，保证了信息的无二义性，从而为信息系统的互操作奠定了基础。

(5) 通信协议的有效性。根据系统网络结构的不同，数据链系统会采用相应的通信协议，保证在不同的网络结构中，通信仍可正常进行。

(6) 信息流程自动化。数据链系统采用统一的格式化信息标准，使战场信息的采集（由传感器完成）、加工（由传感器处理器和指控系统完成）、传输（由数据链终端设备和信道设备完成）、处理（由数据接口设备和指控系统完成）和应用（作战指挥部门和武器平台）能自动完成，无须人工干预，形成信息流程的自动化。从而提高了信息传输的实时性和战场信息利用的有效性，使“传感器到射手”成为现实。

在现代战争中，数据链首先实时、高效、自动的传输、交换和处理信息，大大节约了作战过程中分析、判断和决策的时间，使作战进程加快；其次，数据链将信息覆盖范围变成战场控制范围，扩展了兵力兵器作战的空间性能，使战场的空间朝着纵深化、立体化方向发展；再次，数据链将传统的垂直“烟囱”式信息传输模式变成“扁平”式传输，使树状垂直式作战指挥向网络化扁平式作战指挥转变；另外，数据链嵌入作战平台，摆脱了作战平台过去单一发挥作战效能的状况，紧密的战术链接关系使战术协同能力明显增强，从而为实现真正意义上的一体化联合作战奠定了基础。我们要应对这一变化和挑战，必须摆脱机械化战争时代的观念束缚，树立以网络为基础、以情报为先导、以行动为中心一体化联合作战指挥新理念，在提高信息化条件下作战指挥能力上下功夫、求突破。



作为一种实时化作战指挥手段，数据链运用将对指挥体制、指挥机制、指挥方式和指挥流程产生较大影响。

### 13.1.3 战术数据链的分类

数据链按使用功能可以分为指挥控制数据链、态势分发数据链和武器协同数据链。

按使用对象可以分为通用数据链（如 JIDS、综合链 I 型）和专用数据链（海军数据链、空军数据链、陆军数据链和三军综合数据链等）。

数据链路按工作方式可以分为点对点数据链路和点对多点的网状数据链路。点对点的数  
据链路是在两个单元之间建立一条公共的数据交换信道，它可以很好地把固定或移动的指挥控制  
单元或武器单元连接起来，并且多个点对点数据链路可以通过中继或转发单元连接成网络。多点  
对多点的网状的数据链路通常是建立一个公共的数据信道，为许多网上单元共享。由于信道是公  
用的，所以必须要建立发射协议。网状数据链路适合连接不同的指挥控制单元和武器单元。

按照数据链路的操作方式，大致可以分为 5 类：

（1）点对点全双工操作，如 TADIL B（LINK 1）。其特点是信息是在两个站点之间进行  
全双工传输，主要用于地面指挥和控制单元之间交换监视信息。

（2）点对多点点名呼叫操作，如 TADIL C（LINK 4A），有一个站点作为主控站，其余  
站点是应答站。主控站在控制时隙内发送控制消息帧，而在预留的时隙内等候应答站的应答。  
问到谁，谁应答。正常情况下，不会产生数据碰撞。主要用于对飞机的空中拦截控制和空中  
交通管制等。

（3）点对多点轮询方式，如 TADIL A（LINK 11）。在这种控制方式中，网络中有一控制  
站和多个前哨从属站。网络工作时，网络控制站自动询问每个网络从属站，当从属站识别出  
它的地址被呼叫时，即向网络发出该从属站的数据。所有从属站都被呼叫之后，网络主站发  
出自己的数据，并开始新的循环。该方式主要用于地面和空中单元以及空中单元之间进行监  
视信息的交换。

（4）点对多点时隙分配方式，如 LINK 10。在这种控制方式中，网络中有一主控站和  
多个网络从属站。在一个时间周期内，划分为若干个时隙。网络主控站自动占据第一时隙，  
并首先发送周期起始时间和信息数据，以此同步各网络从属站。各网络从属站在指定分配给  
自己的时隙内，播发消息帧，其余时隙则收听广播或发给自己的信息。该方式主要用于海军  
地面或海面指挥和控制单元之间交换监视信息。

（5）点对多点时分多址方式，如 TADIL J（LINK 16）。在这种控制方式中，网络中是  
没有中心节点的，具有很强的抗毁能力。网络中各个成员依据一定准则分配给其一定数量的  
时隙。在此时隙内，可以发射消息，而在其余时隙内，则接收别站的有用信息。该方式主要  
用于三军联合作战时的信息分发。

## 13.2 战术数据链系统组成与数据传输流程

### 13.2.1 战术数据链系统组成

数据链系统是一种按照规定的消息格式和通信协议，实时传输格式化数字信息的无线传

输系统。通常包含消息标准、通信协议和传输设备 3 个基本要素，设备的组成通常有：战术数据系统（TDS）、系统网络控制器（SNC）、数据终端设备等，具体如图 13-2 所示。

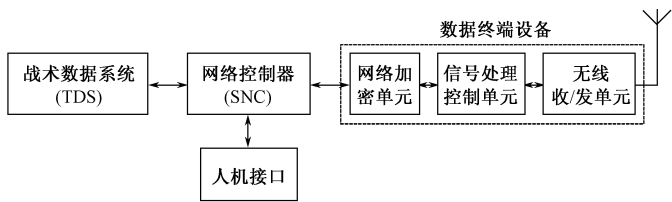


图 13-2 战术数字信息链路组成框图

1. 战术数据系统（TDS）

战术数据系统一般与数据链所在作战单元的主计算机相连，接收各种传感器和操作员发送的数据，按照数据链消息标准转换为格式化信息，并传递给系统网络控制器，同时负责接收网络控制器传送的分组信息。

2. 网络控制器（SNC）

网络管理器主要完成网络管理，提供分组传输任务。它接收来自 TDS 的数据分组，并根据链路号码选择最佳的传输链路进行发送处理，同时接收物理层提交的信号进行分组处理。

3. 人机接口（HMI）

人机接口作为数据链系统的操作界面，使操作人员可以对数据链系统进行具体操作，如参数的加载，系统时钟的注入，任务数据的下载等，还为作战平台与网络的管理和监视提供了交互接口。

4. 数据终端设备

数据终端设备是数据链系统的核心部分和最基本单元，整个数据链网络的通信规程和消息协议都在其中实现。数据终端设备在通信协议的控制下，负责整个数据链的工作，如数据的收发处理，与指挥控制或武器平台进行信息的交换等。数据链终端设备主要由三部分组成：网络加密单元、信号处理与控制单元、无线电收发单元。网络加密单元是为了确保网络数据传输的安全，有时需要对发送的信息进行加密处理。信号处理单元主要完成对数据的纠错编码、对无线电收发设备进行控制等功能。无线电收发设备负责对射频信号的调制解调、接收与发送，为各终端单元提供无线连接的信道。

13.2.2 数据传输流程

数据链系统的工作过程为：首先由作战单元的主任务计算机将本单元欲发送的战术信息，通过 TDS 按照数据链消息标准转换为格式化信息，经接口处理转换后，由端机按照组网通信协议处理，通过传输设备发送。接收方的端机收到信号后，按组网通信协议进行接收处理，再经过接口处理及转换，由 TDS 进行格式化消息的解读，最后送交主任务计算机进行进一步处理和应用，以图形符号等形式显示在作战单元屏幕上。

1. 数据发送过程

数据链系统在工作过程中，数据发送过程如图 13-3 所示。

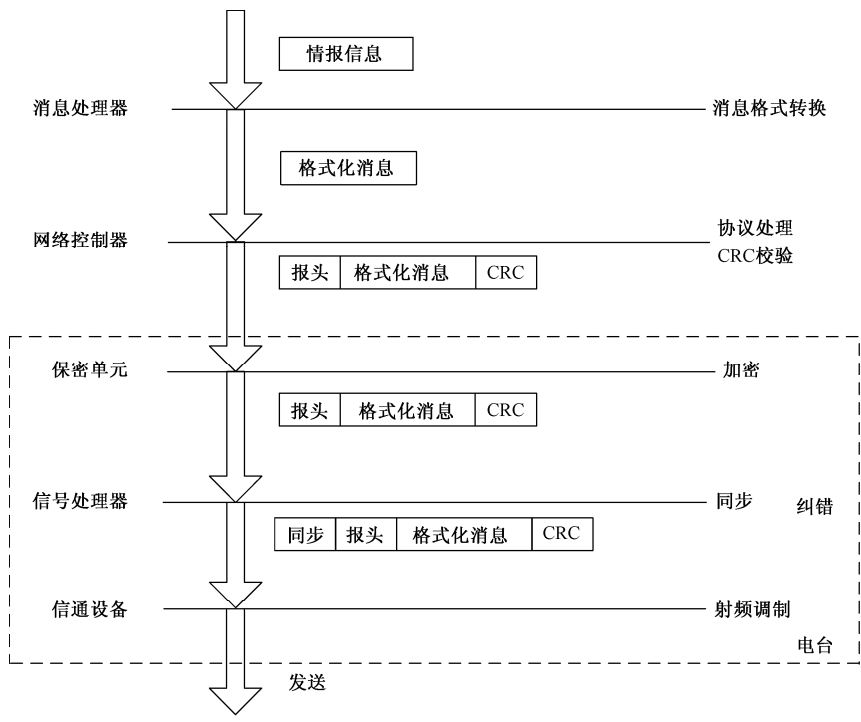


图 13-3 数据发送过程

首先，指控系统、武器系统、传感器系统、显示系统等任务系统将产生的情报信息传送给数据链消息处理器处理。消息处理器对收到的信息进行格式转换、消息标准的翻译等格式化处理，格式化处理后的数据将会存储到发送数据库或发送缓冲区中。等待发送时机到来时，系统网络控制器会从发送数据库或发送缓冲区中取出这些消息数据，进行数据的打包、添加报头（包括链路层报头与网络层报头）和网络控制字，并进行 CRC 校验以提高可靠性。若所发送消息数据需要加密，则在保密单元中进行加密处理，否则直接送入信号处理器，进行纠错编码和添加同步头，然后经调制将信号从无线信道发送出去。其中数据的加密、同步、纠错、调制均在电台完成。

2. 数据接收过程

数据接收过程如图 13-4 所示。

数据的接收过程与发送过程正好相反。从无线信道接收到的信号经解调后送信号处理器进行纠错处理，若处理后的消息数据需要解密则送保密单元进行解密处理，否则送至网络控制器完成协议处理和 CRC 校验，之后将格式化信息送入消息处理器进行拆包和消息格式的转换，最后将处理过的信息送入指控系统、武器系统、传感器系统、综合显示系统等。

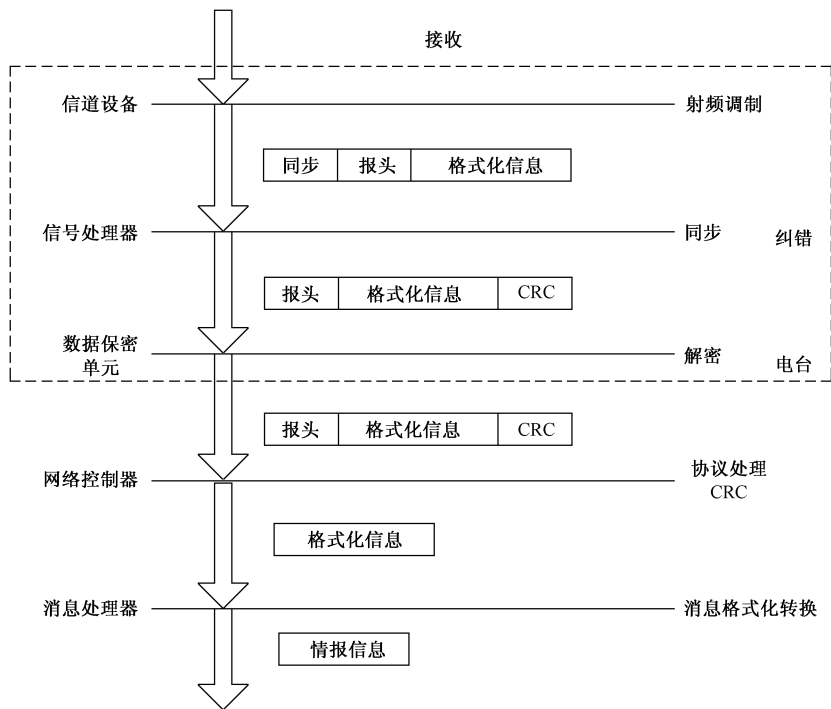


图 13-4 数据接收过程

## 13.3 战术数据链组网协议

数据链是一个标准化的保密数据通信系统，具有 3 个基本的要素，即传输通道、通信协议和标准的格式化消息，其中通信协议包括频率协议、波形协议、链路协议、组网协议、加密协议等。按照分层的思想，战术数据链系统主要工作在物理层、数据链路层和网络层。组网协议是网络层协议的一部分，决定了系统的网络形式，从而决定了系统的循环时间和容量等关键性能指标。战术数据链要负责把实时的战场态势信息分发给指挥系统、武器系统和情报系统，这就要求信息更新的时间间隔足够短，提供的信息足够多，并且要求网络具有较强的抗毁性。因此，组网协议的选择和设计，对于战术数据链系统就显得尤为重要。

### 13.3.1 有中心节点的组网协议

有中心节点的组网协议是指网络节点由中心节点和普通节点构成，中心节点担负着网络运行的管理任务，普通节点参与网络运行的过程中，受中心节点的管理。这种组网方式的优点是组网简单、迅速，普通节点把信息传递给中心节点进行信息融合，不会发生冲突的情况；但它的缺点也很明显，网络抗毁能力差，在作战条件下，中心节点出现故障或被毁则不能管理网络运行，会导致整个网络的瘫痪。Link-4A 和 Link-11 就采用这种组网方式。

#### 1. 指令/回应协议

Link-4A 采用指令/回应协议及时分多路传输，可将一组通信频率分割成可以同时工作的

多个频段，Link-4A 连接两个相互通信单元的方式是分配一系列周期性出现的时隙作为终端接收数据。两个通信单元中，一个作为控制站，另一个作为应答站。控制站在控制时隙内发送控制消息帧，在预留的时间周期内等候应答站的回答，如果在该时间周期内没有收到对当前控制消息的应答，可以在下一个时间周期内发送新的控制消息帧，并在新的周期内接收应答，依照这种方式，直到信息交换完毕。因此，在 Link-4A 的运用上，一个控制平台可在同一频率上同时控制多个不同的作战单元，各个单元在点对点的信道中可以同时发送和接收数据，在不收发送数据时信道呈现静默状态。该数据链每个端机最多可同时在 8 个用户间建立通信网络，实施控制与实时交换战术信息。

## 2. 轮询协议

Link-11 数据链中所有的入网单元（PU）之间按轮询方式工作。在单网中设置一个网控台（NCS），按轮询协议规定，轮流呼叫网络中的各个 PU，PU 只有被 NCS 呼叫到的时候才能发射信号，PU 一旦完成消息发送后就立刻转换成接收模式以备接收下一个周期的呼叫，与此同时，另一个 PU 接收到 NCS 的呼叫后即可向网络发送数据。等所有的 PU 被 NCS 呼叫一次后，NCS 再周而复始地从第一个 PU 开始呼叫。可以看出，在一个网络周期内，每个 PU 都有一次发送数据的机会，采用半双工的通信方式。Link-11 最多可同时容纳 62 个 PU 参与一个网络进行数据通信，但一般使用时 PU 的数量大都限制在 20 个左右，各个 PU 在 NCS 的管理下进行组网通信，一旦 NCS 发生故障或被摧毁，整个网络就会瘫痪，网络的抗毁性较差。

### 13.3.2 无中心节点的组网协议

在无中心节点网络中，各个节点可以互通，组网灵活且网络生存能力强，孤立的节点对网络运行的影响较小。这种组网结构在军事通信中，尤其是在军事战术通信中得到广泛采用。

#### 1. TDMA 组网协议

战术数据链采用 TDMA 组网协议，要在所有成员中指定一个成员作为时间基准，其他成员的时钟与之同步，这个被指定的网络成员便是系统的“中心节点”。在无中心节点网络中，所有网络终端在功能上都是相同的，作为系统的时间基准或导航控制器的网络成员，可以任意指定。在实际操作中，网络终端是排好队的，假如作为时间基准的节点被摧毁之后，接替成员将在某个设定时间内接替执行其任务，所以系统不会因为某个节点故障或被毁而停止工作，产生了很强的抗毁能力。

无中心节点网络中各成员由平等的节点构成，所有成员共同进行管理网络，这就需要在运行之前进行规划或在系统运行时进行初始化，其后自我运行。这种组网协议的缺点是规划起来比较复杂，优点是网络中的各个节点没有主次之分，某一节点是否参与，系统都可以正常运行。

Link-16 数据链最突出的优点就是采用了这种无中心节点结构，每个成员都按照统一的时间基准同步工作。信息被送到网内所有节点都能共享的数据库中，节点之间的信息交换不需要经过中心台的控制和中继，不管哪个节点遭到破坏都不会影响系统正常运行，而且任何一个终端均可起到中继的作用，使系统具有很强的生存能力和抗毁能力。

## 2. 动态 TDMA 组网协议

Link-22 采用动态 TDMA 组网协议,时帧结构和 Link-16 基本相同。Link-22 解决了 Link-11 在每个用户单元和网络控制站之间依靠双向链接维持网络正常运行的问题,提高了数据链系统的可靠性并增加了网络的总体覆盖范围;同时它通过时间分集、频率分集和天线角度分集为用户单元提供一定的冗余度,增强消息传递的可靠性;通过对多个频率上网络运行情况的比较分析选择最优的系统工作频率,提供系统的鲁棒性。Link-22 动态时分多址可以自动实现时隙分配的最优化,支持快速接入、优先级中断和自动实现网络接入延迟等操作,使网络管理更加灵活,建链速度加快。

## 13.4 战术数据链的时隙分配

战术数据链是信息化战争条件下的“兵力倍增器”,其关键技术包括组网协议、信道接入技术、信道编码技术、时隙分配技术等。除时隙分配技术外,其他技术前面已介绍,本节重点讨论数据链的时隙分配。

### 13.4.1 链路终端的入网过程

为了说明数据链的时隙分配技术,首先以 Link-16 为例说明链路终端的入网过程。

Link-16 以 TDMA 方式组网,作为一个基于时间运行的系统,Link-16 要求每个参与 JTIDS 的终端必须获取和保持一个精确的系统时间。作为网络时间基准(NTR)的一个指定节点将统一系统时间,这个角色只可指定给一个节点,一旦工作,该节点将承担精同步任务,并开始广播发送系统时间与入网消息。当网络中某个节点在运行过程中出现故障,对整个网络不造成影响,即使是 NTR 出现故障,整个网络也能在下一个 TDMA 周期恢复正常工作,由预先指定的备用 NTR 代替其工作。其余所有节点的时钟经过入网、粗同步、精同步及同步保持 4 个步骤,完成和保持与 NTR 的同步,形成统一的系统时间。

#### 1. 入网

数据链网络中的各个节点在入网时,搜索并且接收有效的网络时间校准信息,由指派为网络基准(NTR)的某个节点在每个时隙内发送时间校准信息,其余的节点首先随机选择一个时间间隔,在这个时间间隔内扫描该时间校准信息,如果在这个时间间隔内没有接收到时间校准信息,在下一个时隙时,按照算法增加选择的时间间隔,继续监听时间校准信息。

#### 2. 粗同步

根据当前时间的估算值和内部时钟误差的估算值,入网节点从时隙段中选择一个当前还未出现的时隙,开始收听入网消息,如果时间估计正确,就可以接收到入网消息;如果没有收到入网消息,该节点就需要再试一次。如果该节点接收到了入网消息,就用接收到的时间校正节点的系统时间,调整后的系统时间仍然包含由传播时间导致的误差,此时,节点处于粗同步状态。处于粗同步的节点可以接收到网内所有的消息或语音,但只能发送往返计时(RTT),不能发送其他消息。例由于直升机飞行高度的限制,按超短波电台的地空通信距离为 30 km (高

度为 100 m 时) 计算, 则时间校准报文的传输时间 0.1 ms; 通信距离为 120 km, 则时间校准报文的传输时间 0.4 ms, 即经过粗同步后, 各节点与 NTR 的时间误差不大于 0.4 ms。

3. 精同步

网络实现粗同步后, 节点在一个时隙内向作为网络时间基准节点发送一个 RTT 询问消息, 并在同一个时隙内接收 RTT 应答消息。RTT 询问消息既可采用寻址方式, 也可以采用广播方式。询问节点利用报告到达时间与本身测量的应答到达时间得出对本系统时钟的修正, 这样最终达到精同步。网络中各个节点经过以上几步实现同步后, 便可在规划的时隙中实现彼此间的信息收发, 这样数据链完成组网。

4. 同步保持

在数据链系统运行过程中, 通过接收系统时间信息和交换 RTT 保持同步。不管采用哪种方法建立和维护同步, 除 NTR 外所有节点都要生成一个网络时间基准内部时钟模型。如果某个平台丢失了与 NTR 的通信, 它的时钟模型估值将开始漂移。在超视距或干扰的情况下, 根据丧失连通性的时间, 节点将退回到粗同步状态, 重新发送 RTT 并试图再次实现同步; 如果一个平台漂离 NTR 时钟模型太远, 需要重新入网。

13.4.2 时隙及时隙分配

Link-16 数据链把时间划分为循环的时间段, 每 12.8 min 为一个时元 (Epoch), 每个时元又分成若干相等的时间段, 每段 12 s 为一个时帧, 时帧最后细划分成时间长度为 7.812 5 ms 的时隙。时隙以时隙块的形式分配给网络中的节点, 时隙块由等间隔的时隙组合而成。

由此可知, 1 天 24 小时可分为 112.5 个时元, 每个时元又可分为 64 个时帧, 每个时帧可进一步分为 1 536 个时隙。因此, 每个时元包括 98 304 个时隙。一个时元的 98 304 个时隙分为 3 个时隙组, 分别命名为 A 组、B 组和 C 组。每组包括 32 768 个时隙, 编号从 0 到 326 767, 这个编号称为时隙索引号。根据标准, 1 个时元中的时隙是交替命名的。每组的时隙与其他组的时隙交错的排列, 排列顺序为 A-0, B-0, C-0, A-1, B-1, C-1, ..., A-32 767, B-32 767, C-32 767。这个顺序在每个时元中重复出现。时隙结构如图 13-5 所示。

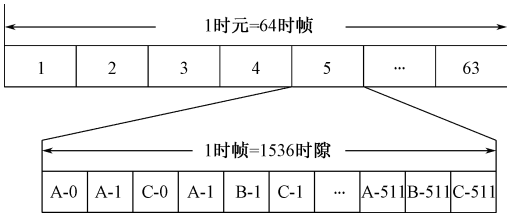


图 13-5 时隙结构示意图

Link-16 数据链周期性的时间单元是时帧。一旦系统开始工作, 时帧重复地出现并一个接一个地传输。时隙是 Link-16 最基本的时间单位, 系统中的每个工作平台在每个时隙中被指定要么接收、要么发送, 时隙既分配给它们所支持的功能, 也分配给它们所参与的平台。如果网络内一个节点分配了一组发送时隙, 并且有信息要发送, 则它将在每个时隙发射一组脉冲, 直到信息全部发出。

13.4.3 时隙传输结构

Link-16 基本传输结构包括抖动、同步、精确定时（TR）、报头、数据和传输保护 6 个部分，如图 13-6 所示。

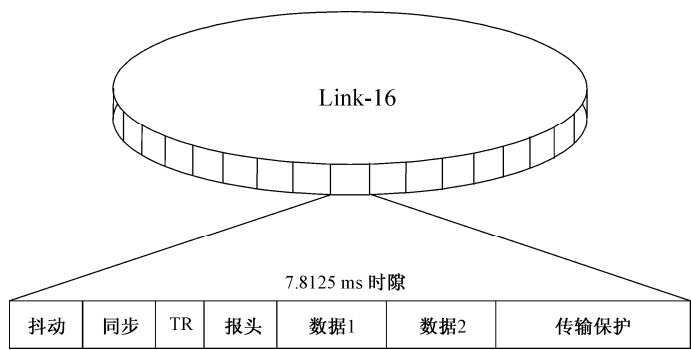


图 13-6 时隙传输结构

- (1) 抖动。从数据传输的有效性来看，应尽量消弱或消除抖动。在 Link-16 的时隙组成中抖动是指一个时隙中传输开始时的一段随机可变的时延，也称为跳时。它通过改变传输的起始时间，提高了 JTIDS 信号的抗干扰能力，但并不是所有的数据封装格式都能使用抖动。
- (2) 同步。同步包含 16 个双脉冲符号，使正在接收的网络单元实现与发射信号的粗同步。同步跳频图案与传输数据的跳频图案不同。在同一个网络中，不同时隙具有不同的同步跳频图案；在同一个时隙中，不同网络具有不同的同步跳频图案。
- (3) TR。精确定时包含 4 个双脉冲符号，其跳频图案是固定的。
- (4) 报头。报头包含 16 个双脉冲符号，提供有关在时隙中传输的时隙类型、中继传输指示符/类型变更、源航迹号、保密数据单元等信息。
- (5) 数据。在时隙中传输的信息包括固定格式、可变格式、自由文本和往返计时 4 种类型。
- (6) 传输保护。传输保护时间用于信号的空中传输，信号传输的时间允许平台之间有 300 海里距离，扩展模式可达到 500 海里。

13.4.4 数据封装结构

根据支持的消息类型、吞吐量需求和抗干扰性能需求，JTIDS 终端能产生 4 种不同的时隙格式：标准双脉冲（STDP）数据封装结构，2 个单脉冲（P2SP）数据封装结构，2 个双脉冲（P2DP）数据封装结构，4 个单脉冲（P4SP）数据封装结构。这些时隙格式在数据脉冲数量、检错纠错编码数量和冗余数量上有所不同。

如图 13-7 所示，4 种消息封装结构都包含 32 个同步头脉冲、8 个精确定时脉冲、32 个报头脉冲和 186 个数据脉冲，两个连续脉冲之间的间隔是 13 μs，STDP 与 P2SP 总的消息传输时间是 2.418 ms+0.936 ms=3.354 ms，消息封装结构都包含抖动；P2DP 与 P4SP 总的消息传输时间是 2.418 ms+2.418 ms+0.936 ms=5.772 ms，消息封装结构都不包含抖动。STDP 在时



隙的 2 个脉冲中传输相同的数据，提供了低吞吐量和强抗干扰能力；P4SP 在每个脉冲信号中发送不同的数据，使吞吐量加倍；P2DP 加倍封装增加了吞吐量，传输冗余脉冲恢复了一些抗干扰能力，但以损失抖动为代价；P4SP 消息封装结构放弃了抖动和脉冲冗余，降低了抗干扰性能，但获得了最大数据容量。

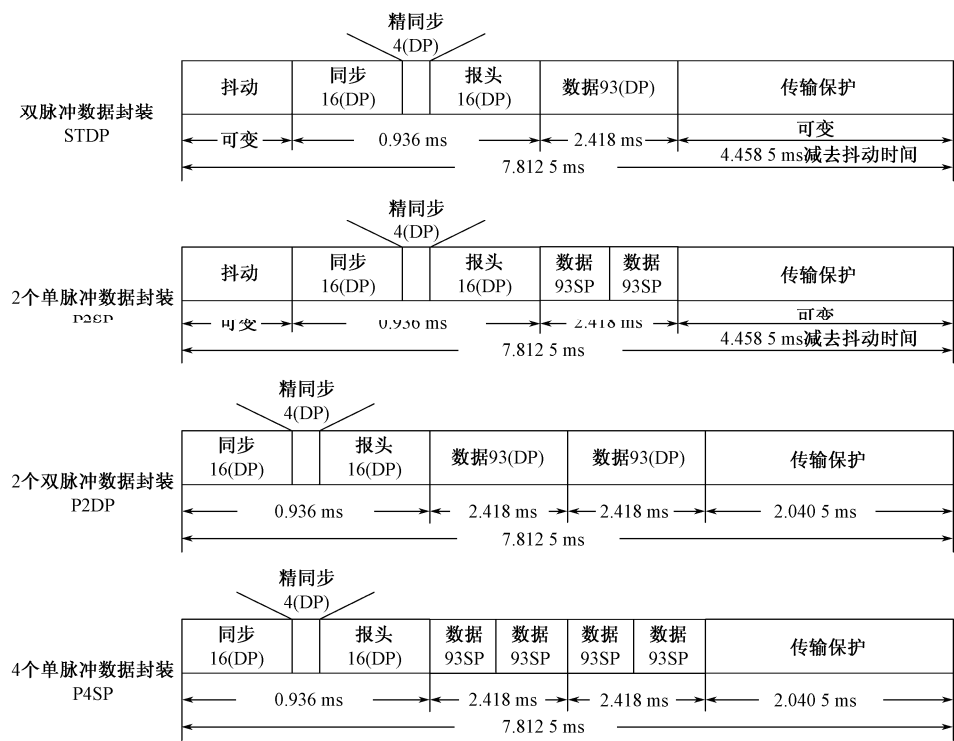


图 13-7 数据封装结构

## 思考与练习

- 13-1 何谓战术数据链？试说明战术数据链的基本三大要素。
- 13-2 试画出战术数据链系统组成框图，并简要说明各部分的作用。
- 13-3 简述战术链路终端的入网过程。

# 第 14 章 自由空间激光通信技术

自由空间激光通信又称无线光通信或自由空间光通信（Free Space Optical Communication, FSO），是光纤通信和无线通信相结合的产物，它是指以  $10^5$  GHz 以上的激光作为信息载体，不使用光纤等有线信道的传输介质，在宇宙或大气中传递信息的一种通信技术。

从物理机制来看，激光通信与无线电通信基本相似，在发送端用激光器发出的激光作为载波。数据信号调制在激光光波上，通过发射望远镜（也称发射天线）发送出去，在自由空间（宇宙或大气）媒质中传播。在接收端，接收望远镜（也称接收天线）将激光信号按发送端的逆方向转化为数据信号。

从物理学的观点来看，激光是一种光波，也具有电磁波的性质。然而，激光与一般的无线电波又有明显的不同，激光的频率为几亿兆，是微波（超高频电磁波）频率的 10 万倍以上。由波长与波速及频率的关系式可知，激光的波长非常短，所以其波动性远比无线电波差，相反，激光却具有奇特的粒子性，因而使它在军事通信和民用通信中成为引人注目的“后起之秀”。

## 14.1 自由空间光通信系统的基本组成

自由空间光通信利用激光为载波，在空间进行信息和数据的传输。自由空间光通信系统可应用于海岸与海岸之间、岛屿与岛屿之间、舰船之间、城市高层设施之间，以及卫星与卫星、卫星与地面之间的通信等。

自由空间光-光通信是一个复杂系统，根据实际工程的要求，可以将系统从逻辑上而不是物理上分成两个适度相关的子系统：通信子系统；捕获、对准和跟踪（Acquisition Tracking Pointing, ATP）子系统。其系统组成主要包括光源系统，发射与接收子系统，对准、捕获、跟踪子系统，以及其他一些辅助系统等，如图 14-1 所示。

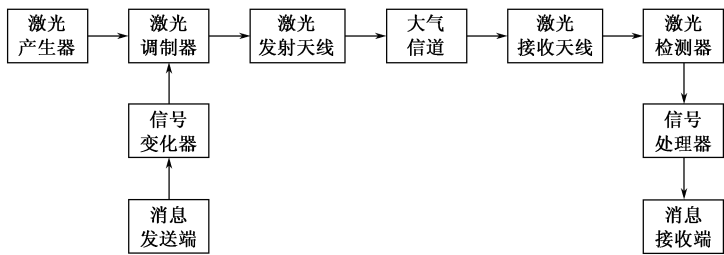


图 14-1 自由空间光通信系统基本组成

### 1. 光源部分

由于大气性能的影响，进行激光空间通信时必须考虑激光的功率和传输频率等特性。目前应用于通信系统的激光器有半导体激光器、固体激光器如 YAG 激光器、气体激光器等。半导体激光器可以产生波长为  $1.06\text{ }\mu\text{m}$  的激光，这是大气的一个低损耗窗口，因而可以有效

减少激光传输中的衰减。此外，半导体激光器具有体积小、转换效率高、成本低并可直接调制的特点，故实际通信系统一般均采用半导体激光器做光源。

## 2. 空间光发射与接收部分

发射与接收是空间激光通信系统中最重要的部分，包括调制（解调）器、光发射（接收）天线及探测放大等装置。发射机主要是完成信息对光源的调制，并通过光学天线发出承载信息的激光，而接收机用来接收信号光、进行光电转换，通过放大、解调等处理完成通信任务。

## 3. 对准、捕获、跟踪子系统

激光空间通信与微波等无线通信方式不同，是近似的点对点通信，所以发送光必须准确地到达接收机探测器上。

用调整发送的激光瞄准一个特定方向的过程称之为对准（Pointing）；确定入射光到达目标方向的接收过程称之为捕获（Acquisition）；调整整个通信过程中对准和捕获的动态变化过程称之为跟踪（Tracking）。

APT 子系统的功能是进行空间激光通信链路的建立与保持，即 ATP 子系统对运动的目标进行捕获、跟踪和对准。捕获是指在预计的目标可能存在的区域（即不确定区域）对目标的识别。跟踪是仪器视轴跟随目标视轴的晃动。瞄准是仪器视轴对目标视轴的平均取向。针对于空间激光通信，由于光束宽度非常窄，光束的跟踪和瞄准要求精度很高，卫星在轨道上空间位置的不确定性远远大于跟踪和瞄准误差，使得短时间内的概率捕获也成为是一个非常困难的问题。ATP 技术难度很大，是限制星间激光通信的“瓶颈”之一。

在进行空间激光通信时，要尽量减少信道中的衰减和干扰因素的影响，同时要求通信具有较高的传输码率和较高的保密性能。由于自由空间信道具有随机的不稳定性，所以通信系统必须具有实时调整的 ATP 来适应不断变化的条件。

## 4. 空间光通信子系统

光通信子系统的作用是将进行卫星之间的数据传输与信息交换。通信双方通过 ATP 子系统相互跟踪锁定，进入跟踪过程后，才开始进行激光通信。在 APT 子系统保证跟瞄精度的条件下，通信子系统一方面要完成高码速率的电信号驱动与调制以及大功率的光信号发射，另一方面实现极高灵敏度的微弱光信号的检测、接收与基带信号的再生。

通信子系统分成几个相对独立的模块，主要包括光源调制驱动模块、发射光功率放大模块、光准直模块、空间光—光纤耦合模块、前置光放大及信号处理模块等。

光源调制驱动模块的作用是将电信号转变成光脉冲信号，由于在接收端通常采用窄带滤光器以滤去背景杂光，所以要求光源发出的激光频率稳定。调制驱动的过程中应保证激光器的张弛震荡小。光功率放大模块的主要作用是实现发送功率的助推，将信号光功率放大到 W 量级，然后经过天线发射出去。

通常星间激光通信要求发射光束的宽度达到  $10\ \mu\text{rad}$  左右，因此光依靠光学天线的准直作用是远远不够的，光准直模块的作用就是对信号光进行预准直，从而使得从发射天线射出光束达到系统要求。对于采用前置光放大器以提高接收机灵敏度的系统，需要采用空间光——光线的耦合模块将天线收集的信号光注入到光纤放大器中加以放大。如何通过光学天线接收到的空间光高效地耦合进入单模光纤是一个需要解决的重要问题。前置放大模块位于

光接收与信号处理模块之前，该放大模块的主要作用是对经空间传输衰减、光学天线接收和空间光——光纤耦合后的微小光信号进行预放大，以抑制接收机内的噪声，从而提高光接收机的接收灵敏度。

采用高灵敏度的探测器是提高接收灵敏度的最直接、最简单和最有效的方法。然而，探测器灵敏度受器件水平的限制。

电转换模块及信号处理模块用来完成光/电转换，实现数据再生任务。对由接收端前面采用 EDFA 前置光纤放大器的情况，其输出的光信号功率已经足够强，因此对光电转换模块及信号处理模块的接收灵敏度要求不高。但是对于采用光电探测器直接接收的情况，其接收灵敏度应尽可能的提高性能。星间激光通信信标光光束和信号光的光束相对于微波通信要窄得多，光束窄使得星间激光通信具有很多优势，比如：加强保密性，降低发射功率，减小系统及元器件的尺寸、重量功、耗等，但同时也带来一些困难。

## 14.2 FSO 的技术特点与优势

与短波、超短波、微波、毫米波、有线、光纤等通信系统相比，宇宙和大气空间的激光通信主要有以下技术特点与优势：

### 1. 不怕核辐射、不怕侦听、不怕干扰

一般无线电通信方式在核武器爆炸瞬间都要中断一阵子，唯独激光通信不受其害，这是因为核武器爆炸所产生的强电磁脉冲，并没有激光的频率高，所以干扰不了激光。同时，核武器爆炸所产生的强烈光辐射，属于非相干光之类，尽管能量很强，但是不集中，向四面八方发散，即使能够钻进激光接收机，其强度相比也是微乎其微，除了会增加点噪声外，无其他影响。

另外，由于激光频率极高，光束极细，方向性又好，所以用当今的设备无法侦听和截获，用电磁波更难干扰。

### 2. 信息容量大

理论上，无线光通信的传输带宽与光纤通信的传输带宽相同，具有频带宽、信息容量大的优势。目前实验室中研制的 FSO 设备甚至能达到 160 Gb/s，因此也有人把无线光通信技术称为虚拟光纤通信技术。

### 3. 设备重量轻、尺寸小、功耗小

由于激光波长短，在同样功能情况下，激光收发终端的尺寸比微波、毫米波通信天线尺寸要小，只有 10~30 cm。另外，FSO 所使用的固体和半导体激光器，发光效率很高，功耗却很小，不需要有庞大的能源供给设施，因此系统设备极易小型化，无论由机载、舰载、还是卫星承载都十分理想。

### 4. 安全保密性高

自由空间光通信的安全保密性较好，因为红外激光的波束窄，且不可见，很难在空中发现其通信链路。同时，激光束定向性好，能量集中，不向空中其他方向产生辐射，激光光束

的发散角通常都在毫弧度，甚至微弧度量级，要求通信两点间必须无阻碍，任何对光束的遮挡都将对通信造成影响。如果想截取，一般需要在链路中插入接收机，并对准发射机，这是很难做到的，即使被截接，还需要知道信号的解密办法，另外，用户也会即时发现，因为链路被中断了，因此说 FSO 比通常的无线系统安全得多。

5. 透明传送，适用任何通信协议

FSO 是物理层传输设备，任何传输协议均可容易的迭加上去，对语音、数据、图像等业务，以及任何通信协议，如 SONET、SONET/SDH、ATM、FD-DI、以太网、快速以太网等，都可以实现透明传送。

6. 频谱资源丰富，不需要申请频率使用权

目前，无线通信方式的频率几乎被瓜分完毕，发展余地已所剩无几，而目前 FSO 设备，采用红外光频带传输，且设备间无射频信号干扰，所以在全世界范围内均不受管制，故无须申请频率使用许可证，在有效通信距离和宽带等方面还蕴藏着巨大的发展潜力，为无线光通信技术的灵活应用提供了有利条件，这是一般微波通信和无线通信无法比拟的。

7. 地表空间商用通信时，建网快速、造价低

地面空间进行商用激光通信时，因为无须进行昂贵的管道工程铺设，不存在埋设光纤和等待各种手续上的问题，只须在通信点上进行设备安装，工程建设以小时或天为计量单位，故建网快速，对于重新撤换部署也很方便、容易，其造价约为光纤通信工程的 1/5。

14.3 最新国内外研发进展与应用情况

14.3.1 国外研究的最新进展与应用情况

1. 美国进行“转型通信卫星”的自由空间激光通信侧试

美国空军将自由空间激光通信视为未来能提供高带宽、基于互联网协议的卫星通信的转型卫星通信系统所需的关键风险技术。

2006 年 5 月，美国进行“转型通信卫星”的自由空间激光通信侧试，在洛克希德-马丁空间系统公司领导小组的领导下，负责向美国空军提供 TSAT 空间部分的诺斯罗普-格鲁曼空间技术公司，对符合美国政府新规定的激光通信互通标准的高数据率激光通信进行了测试，公司测试了高数据率激光通信定位、探测和跟踪另一个激光终端，并在航天器出现颤动时保持跟踪的能力，试验结果表明，硬件能以 10 Gb/s 以及 40 Gb/s 的数据率运行。8 月，波音公司在一个模拟的太空环境中，成功演示转型卫星通信系统利用激光束在卫星间的通信能力，这次演示是第 3 次具有里程碑意义的激光通信演示试验，在演示中，激光通信传输速度为 10~40 Gb/s。

在未来几十年中它将成为对美国安全起着至关重要作用的一项关键技术，也成为美国国防部转型通信卫星的一大要素。

## 2. 日、德成功完成卫星与地面站间的激光通信试验

日本继“光学轨道通信试验卫星”(OICETS, 又名 Kirari, “闪光”)与欧空局“先进中继及技术任务”卫星(ARTEMIS)实现了世界上首次卫星之间的双向激光通信后, 2006 年 3 月 22~31 日, 日本国家信息通信技术研究所有(NICT)光学地面站与日本航宇局之间进行了光学通信试验, 这是世界上首次成功进行的低地球轨道卫星与地面站间的光学通信试验。6 月 7 日, 日本航宇局利用激光束成功进行了“闪光”卫星与德国航宇中心光学地面站 OGS-OP 间的光学通信试验。这次使用的德国光学地面互通为移动地面互通, 因而这次成功意味着利用卫星与移动光学地面互通建立灵活的光学通信网络的可能性。

## 3. 法国用激光首次在卫星与飞机之间建立通信链路

2006 年 12 月 13 日, 法国国防部利用激光束首次在卫星与飞机之间成功建立了通信链路。“先进中继及技术任务”卫星与飞行在万米高空的“神秘”20 (Mystere20) 商用飞机之间建立的高速通信链路时, 在不足 1 s 内就可完成数据传输率为 50 Mb/s 的激光链接建立。

## 4. 国外商用研究与应用情况

美国是世界上开展军用和商用空间光通信最早的国家之一, 2001 年 2 月由美国的一些光无线设备制造商联合电信运营商, 成立了空间光通信联盟 (FSO Alliance)。现在, 美国 AstroTerra 公司已研制出多种光无线接入产品, 如 TereScope 系列提供负载光层 IP 业务功能以及 1.5 Mb/s (4.1 km) ~ 1.25 Gb/s (500 m) 的速率。美国 LightPointe 公司将自由空间光学技术用于研究、设计和制造电信公司等级的光传输设备, 目前 LightPointe 提供的系统以超快的带宽速度提供安全可靠的无线传输, 速度最高可达 2.5 Gb/s, 产品适应性强, 可解决城市地区的连接问题。

日本 Canon 利用在光学系统方面的优势, 也涉足自由空间光通信系统的领域。主要产品为 Canobeam, 可以桥接最远 2 公里的距离, 支持从 25 Mb/s 到 1.25 Gb/s 的各种数据速率。特点是具有自动跟踪系统, 调整探测器件的位置以检测激光束的光轴, 可以不因建筑物的摆动和震动而使传输中断。

### 14.3.2 国内研发的进展与应用情况

国内已有桂林 34 所、清华同方股份有限公司、中科院成都光电技术研究所、深圳飞通光电股份有限公司、上海光机所等研究所与公司从事 FSO 研发, 虽和国外相比起步晚, 但在发射功率、接收灵敏度、捕获和瞄准、机械稳定性等关键技术方面已取得明显进步。

清华同方最早推出了商用自由空间通信产品 OWLink E100, 提供了可以无缝集成到已有以太网的 10/100 M 以太网接口, 并采用自主专利技术的电路设计方案和软件控制技术, 实现自动跟踪和自动搜索, 使系统能快速精确对准和保持光链路的连接。

深圳飞通开发出光收发模块加上 EDFA 系统方式, 其速率有 155 Mb/s、622 Mb/s 以及 2.5 Gb/s 几种, 通信距离最远可达 4 km。

在实际应用方面, 长城宽带网络公司 (GWBN) 已选用 Terabeam 的 FSO 产品来拓展其在中国的 15 个城市的宽带网络; 上海铁通选用 50 套 FSO 产品, 安装在陆家嘴金融区; 2008 年,

FSO 成为北京奥运通信设备，并作为光纤与 3G 网络的有力补充，圆满地完成了奥运通信任务。

## 14.4 存在的问题和解决思路

### 1. 存在的问题

由于激光在接近真空的空间环境中传播时几乎不衰减，采用小功率激光器就能进行远距离通信，卫星和卫星间的光通信的关键技术已经基本解决，美中不足之处是激光在地球表面的大气中传输时，受云、雾、雨、雪、尘埃的影响，产生散射和吸收损耗，衰减严重。其中，雾对 FSO 的影响最大，这是由于 FSO 的波长接近雾粒，能量被吸收；同时，雾粒呈现出棱镜的作用，使激光产生衍射的结果。恶劣天气状况下，每公里造成的损耗可达 50~300 dB，使得 FSO 系统的有效工作距离不到 50 m。另外，激光光束只能直线传播，不能绕过障碍物，应用范围也受到一定限制。

目前，FSO 系统有 780~850 nm 和 1 520~1 600 nm 两个波长传输窗口。1 520~1 600 nm 的大气衰减作用小，射到人眼时，大部分能量都会被眼角膜吸收，不会对视网膜造成影响，允许的功率可以比 780~850 nm 的设备高两个等级，可达 10 mW，有利于提高传输距离，并在一定程度上抵消恶劣气候给传输带来的影响，提供距离可达 4 km 的点对点传输和高达 2.5 Gb/s 的商业应用，这个通信距离和速率与光纤通信相比仍很低，且不能全天候使用。

### 2. 解决思路

解决此类问题，许多研究机构和厂家提出了各具特色的设想和产品，我们认为可行方案如下。

#### 1) FSO 与微波系统 (RF) 互备份

解决问题的一种方法是使用不需要许可证的 W 频段微波 (2.4 GHz, 5.8 GHz, 60 GHz 等) 作为备份节点，这是因为微波对雾的抗力强一些，而在实际使用中只有很少的时间需要切换到微波上。

现在，已有厂商推出了工作在 60 GHz 频段，数据传输速率高达 1.25 Gb/s，和 FSO 技术结合的复合型产品，在极端天气出现时，设备可自动在 FSO 和 W 频段模式间切换。在一般天气里，两类设备同时运行，其中一台作为冗余备份，这一方面可将可靠性从单台的 99 % 提高到 99.99 %，另一方面还可防止意外的光线遮挡。

#### 2) 扩束法和多束法

扩束法是展宽激光的发射光束，但扩束法降低了接收端的光斑能量密度，传输距离和速度受到影响和限制。多束法利用多个激光器和发射镜同时发射激光束，每个光束都以相同的发射角发射，在接收处就得到一个大而相互重叠的激光光斑。从而提高了接收端的能量密度，也扩大了可接收面积。

#### 3) 光束对准和动态跟踪

自由空间光通信系统通常以点到点的方式进行传输，由于激光光束较细，为了保证正常、

稳定的通信，通信的两端点必须保证稳定的对准状态，应采用 ATP (Acquisition, Tracking and Pointing) 技术，即捕获、跟踪、瞄准技术。捕获（粗跟踪）是指在较大视场范围内将 FSO 的两个通信端点对准在一定范围内，以便自动跟踪系统可以正常工作，可采用 GPS 技术全自动完成。跟踪、瞄准（精跟踪）系统，可采用四象限红外探测器 QD 或 Q-APD 高灵敏度位置传感器来实现，并配以相应的电子学伺服控制系统。

#### 4) 开发利用 T-RAY 频段

无线电波的频率比光低很多，可以轻松穿过浓雾；但在普通的无线频谱内，不可避免地又会遇到普通无线电技术所无法回避的一系列麻烦。因此，可考虑传统无线频谱之外的无线频谱，目前，还未被开发利用的频谱基本集中在 T-RAY 频段，这一频段与红外光极其接近。工作在这一频段的 FSO 设备可以将无线电波和可见光的优点结合起来，在雪、雨、雾天气中也可以高速传输数据。

总之，FSO 将来的研究将集中在增加传输容量、延长传输距离、自动方向对准、降低设备成本等方面。如果这些问题能得到有效解决和性能提升，那么 FSO 将一定发挥巨大潜能和优势。

## 思考与练习

- 14-1 简要说明 FSO 的技术特点与优势。
- 14-2 试画出自由空间光通信系统基本组成，并简要说明各部分的作用。
- 14-3 说明自由空间光通信系统中 ATP 的含义。



## 参 考 文 献

- [1] Richard A Poisel. 现代通信干扰原理与技术. 陈鼎鼎, 等译. 北京: 电子工业出版社, 2005.
- [2] 林家薇, 杜思深. 军事通信技术基础[M]. 西安: 西安电子科技大学出版社, 2001.
- [3] 王建平, 等. 通信原理[M]. 北京: 人民邮电出版社, 2007.
- [4] 吴资玉, 等. 通信原理[M]. 北京: 电子工业出版社, 2008.
- [5] 达新宇, 等. 通信原理教程(第二版)[M]. 北京: 北京邮电大学出版社, 2009.
- [6] 张德纯, 等. 现代通信理论与技术导论[M]. 西安: 西安电子科技大学出版社, 2004.
- [7] 杜思深. 综合布线[M]. 北京: 清华大学出版社, 2010.
- [8] 杜思深. 通信工程设计与案例[M]. 北京: 电子工业出版社, 2009.
- [9] 夏靖波, 符艳军. 网络安全技术[M]. 西安: 空军工程大学电讯工程学院, 2003.
- [10] 陈火旺, 等. 网络安全技术[M]. 长沙: 中南大学出版社, 2005.
- [11] 王育民, 刘建伟. 通信网的理论与安全[M]. 西安: 西安电子科技大学出版, 1999.
- [12] 边居廉, 等. 通信保密技术[M]. 北京: 国防工业出版社, 2003.
- [13] 戴逸民, 等. 通信网的安全理论与技术[M]. 北京: 清华大学出版社, 2006.
- [14] 张红旗, 等. 信息网络安全[M]. 北京: 清华大学出版社, 2003.
- [15] 蒋睿, 等. 网络信息安全理论与技术[M]. 武汉: 华中科技大学出版社, 2007.
- [16] 姚富强. 通信抗干扰工程与实践[M]. 北京: 电子工业出版社, 2008.
- [16] 于全. 战术通信理论与技术[M]. 北京: 电子工业出版社, 2009.
- [16] 李文海, 等. 数字通信原理(第二版)[M]. 北京: 人民邮电出版社, 2007.